Facial identification from online images for use in the prevention of child trafficking and exploitation

Ching Yiu Jessica Liu

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

September, 2018

Acknowledgement

I would first like to express my sincere gratitude to my director of studies Prof. Wilkinson for her continuous support throughout my academic life since my undergraduate degree. Her guidance and knowledge has widened my perspective on the possibilities of the collaboration between Art and Science. Thank you for all the opportunities throughout the years, especially in doing this PhD and in being a part of the Face Lab team.

I would like to thank the rest of my supervisory team Dr. Martin Hanneghan and Dr. Sud Sudirman from the Department of Computer Science for their insightful comments, hard questions and encouragement. They have helped me to understand the 'best of both worlds' throughout the journey of this interdisciplinary research. In addition, I would like to thank the Faculty of Engineering and Technology for funding this PhD and making this possible.

My sincere thanks go to Face Lab, especially Mark Roughley, Dr. Sarah Shrimpton and Kathryn Smith, for the stimulating discussions, collaboration on projects, food and all the fun we have had in the last four years. They have helped me settle in Liverpool and taught me so much through the diversity of their skill sets. It has been a wonderful journey and experience working with the team. I would also like to acknowledge the Mathematics and Statistics Department, especially Dr. Carole Turner in helping me to understand and visualise my data.

I thank my friend and flatmate Dr. Akilah Jardine for the inspiring discussions which often went on for hours. Her helpful PhD tips, mental support, and laughter has truly made this PhD journey a lot smoother. Thanks also go to the Liverpool University Sub-Aqua Club in providing the opportunity to explore another world. Special Thanks goes to Jason for keeping me sane through the last stage of my PhD.

Last but not the least, I would like to express my deepest gratitude to my family for their loving support and sacrifices. More than a decade of education away from home has not been easy and the opportunities they have given me are unimaginable.

Without the precious support from everyone, this journey would not have been possible.

Table of Contents

List of Figures

List of Tables

Abstract

Every year, an estimated 1.2 million children are trafficked (International Labour Office, 2002). The National Center for Missing & Exploited Children (NCMEC) received a 432% increase in child sexual abuse images for the purposes of identification between 2005 and 2009 (U.S. Department of Justice, 2010), and they assisted in the identification of 2,589 victims related to indecent images of children in 2015 (NCMEC, 2015a). In relation to the vast number of images received, machine-based facial recognition could help law enforcement and other organisations to match faces more efficiently. The use of facial recognition technology has become more popular within our society, but where rapid juvenile growth changes facial features recognition is challenging, especially for children under 15 years of age with changes to the outer contour of the face (Ramanathan and Chellappa, 2006). The challenge not only relates to the growth of the child's face, but also relates to face recognition in the wild with unconstrained images.

This study aims to provide an open-access database of facial images, documenting the different stages of facial growth from numerous individuals from birth to 19 years of age. There are currently very limited longitudinal databases available for the research community, and the collection of this database will benefit all researchers who wish to study age progression and facial growth.

Ferguson (2015) suggested that facial recognition algorithms can perform better than humans in the identification of faces of children. Experiment 1 of this research takes a further step to explore how the difference in age group and age gap can affect the recognition rate using various facial recognition software, and explores the possibilities of group tagging. Results indicated that the use of multiple images is beneficial for the facial identification of children.

Experiment 2 explores whether age progression work could further improve the recognition rate of juvenile faces. This study documents the workflow of a new method for digital manual age progression using a combination of previously published methods. The proposed age progression method for children recorded satisfactory levels of repeatability with facial measurements at the Nasion (n) and Trichion (tr) showing the most inaccuracy.

No previous studies have tested how different conditions (i.e. blurring, resolution reduction, cropping and black and white) can affect machine-based facial recognition nor have they explored the relationship between age progression images and facial recognition software. The study found that reduction of the resolution of an age progression image improves automated facial recognition for juvenile identification, and manual age progressions are no more useful than the original image for facial identification of missing children.

The outcome of this research directly benefits those who practice facial identification in relation to children, especially for age progression casework.

1. Introduction

This research was supervised jointly from Face Lab at Liverpool School of Art & Design and the LJMU Department of Computer Science. This is *not* research in computer science, but a multidisciplinary project involving areas relating to humanitarianism, facial anthropology, art, and science. A background in forensic anthropology provides basic research skills to utilise existing technology and methodology for analysis. This project applies knowledge in computer science as building blocks to answer questions relating to the facial recognition of children.

Chapter 1 is divided into three sections. Chapter 1.1 addresses the scale of the problem and explores the motivations behind this research in order to identify the research gap and explore related research in this field. Chapter 1.2 addresses published literature and how the results of this research have been applied in forensic casework. Chapter 1.3 addresses research in age progression.

1.1 The scale of child trafficking and exploitation

Child trafficking is the *"recruitment, transportation, transfer, harbouring or receipt of children for the purpose of exploitation*", and this applies to victims under eighteen years of age (UNODC, 2004). Under the Palermo protocol (ECPAT UK, 2015; OHCHR, 2000), because a child is unable to give consent to being exploited, only movement and exploitation are required in order to be defined as trafficking. Child trafficking is recognised by the United Nations as one of the major violations of human rights (UNODC, 2004), it is a form of child abuse and modern day slavery (CEOP, 2011) affecting children locally and on a global scale. Between 2010 and 2012, the number of victims identified in child trafficking from the 80 UN countries was around 10,000 (UNODC, 2014), but the official figures were thought to be the 'tip of the iceberg', where the more realistic number of global victims is unknown. The reference figure provided by the International Labour Office (2002) suggested that every year, an estimated 1.2 million children are trafficked. Since the different forms of trafficking are often analysed as separate entities, there are no published up-to-date figures for globally trafficked children.

Three main stages of child trafficking are recognised (Figure 1) by UNODC:

1. Recruitment; where the child is first enlisted by the trafficker
2. Movement; where the child is relocated locally, regionally or even internationally
3. Exploitation; where children are traded for purposes such as labour, sexual abuse, crime, armed conflict, organ transfer, child begging, adoption, and benefit fraud etc. (CEOP, 2010; International Labour Office, 2008).

The diagram originally presented here cannot be made freely available via LJMU E-Theses Collection because of Copyright. The image was sourced at UNODC (2015). *Human Trafficking*. [Online] Available at: http://www.unodc.org/unodc/en/human-trafficking/what-is-human-trafficking.html [Accessed 3.7.15].

Figure 1: Elements of human trafficking (UNODC, 2015)

Children are coerced into trafficking for many reasons: pressure to help support their family; lured by the relationship of trust; promised a better life or income by moving away; or be trafficked alongside their family (International Labour Office, 2008). Kidnapped or abducted children were also exploited, but this is much rarer than other forms of trafficking (International Labour Office, 2008).

The most common purpose of human trafficking was forced labour and sexual exploitation (UNODC, 2016, 2014) (Figure 2), and the International Labour Office (2002) estimated that 5.7 million children were forced into bonded labour, with 5.5 million concentrated within the Asia-Pacific area. Children may be exploited to work in agriculture, mining, construction, factories, entertainments, nail bars, hospitality or domestic servitude etc. (CEOP, 2011; International Labour Office, 2008).

The diagram originally presented here cannot be made freely available via LJMU E-Theses Collection because of Copyright. The image was sourced at UNODC (2016) *Global Report on Trafficking in Persons*. [online] Available at: https://www.unodc.org/unodc/en/data-and-analysis/glotip.html [Accessed 8.2.2018]

Figure 2: Forms of exploitation in 2014 (UNODC, 2016)

In 2005, the European Commission estimated that there were one million child sexual abuse images online, and this number increases by half a million each year with 70% of the victims being under 10 years of age (European Commission, 2015). The National Center for Missing & Exploited Children (NCMEC) received a 432% increase in child sexual abuse images for purpose of identification between 2005 and 2009 (U.S. Department of Justice, 2010). Along with the rising concerns in relation to child sex tourism, the U.S. Department of State (2007) estimated that more than 2 million children were sexually exploited every year on a global scale, with an estimated 1 million children forced to sell their bodies every day within the sex industry. Southeast Asia, Thailand and Cambodia, in particular, were popular destinations for sex tourism (Rafferty, 2007).

Commercial sexual exploitation of children (CSEC) is a term used to describe the combination of child prostitution, child sexual abuse materials and child sex tourism (ECPAT International, 2015). The global figures reported by various organisations related to CSEC were criticised as being inaccurate, where the recording methods used were often unstandardised and inaccurate based on very limited resources. These issues were often focused separately and it was described as a "hidden population" where the representative sample was very limited (ECPAT International, 2015). However, in 2017, 78,589 web pages were reported to the Internet Watch Foundation (IWF) and confirmed as containing child sexual abuse material (CSAM), 8,974 were commercial in nature (IWF, 2018).

CEOP provided child trafficking data (Figure 3) and suggested that the UK was a significant transit point and destination for child trafficking from regions such as Africa, Asia and Eastern Europe (CEOP, 2012). Most victims were between 14-17 years old, and victims from Africa and Eastern Europe were mostly female, whereas victims from Asia were mostly male (CEOP, 2012). Victims from Africa were mostly traded for sexual exploitation, victims from Eastern Europe for benefit fraud and criminal exploitation, and victims from Asia were mainly traded for labour exploitation, with many identified in the drug trades, such as cannabis cultivation (CEOP, 2012; CEOP, 2011).

Figure 3: Purpose and origin of child exploitation within the UK (CEOP, 2011)

Child trafficking and exploitation remain a social taboo, often unspoken and under-reported. However, these official figures may not correlate with the number of missing persons, as children can be sold into abuse by their families due to poverty (U.S. Department of Justice, 2010). The Chinese state media reports have estimated that 58 million children were abandoned by their migrant-worker parents (U.S. Department of State, 2012), therefore children in these circumstances may never be reported as missing.

1.1.1 **Technology and Child Trafficking**

Online child sexual exploitation (OCSE) is a rising problem. In 2012, CEOP reported 1,145 cases of OCSE within the UK. Importantly, approximately 5% of children suffered from contact sexual abuse, meaning 10,000 new victims in the UK every year (CEOP, 2013). As technology advances, storage and distribution of indecent images of children becomes easier through the hidden internet, webmail, social networks, file hosts, peer to peer file sharing, or live video streaming (CEOP, 2013). IWF reported a rise in the sharing of child sexual abuse material (CSAM) via redirecting hacked websites, and the removal of these newly identified commercial CSAM websites could be challenging (Smith, 2014a;b).

Within 2012, 50,000 individuals within the UK alone were involved in sharing and downloading indecent images of children, this involved 70,000 still and moving indecent images of children, which was a two-fold increase compared to 2011 (CEOP, 2013). With the 432% increase in child sexual abuse images reported by NCMEC, existing databases managed by INTERPOL contained more than 500,000 indecent images of children within the International Child Sexual Exploitation (ICSE) Database, (INTERPOL, 2011; Wei, 2012). Materials retrieved from online sources can be a useful tool in finding the missing child and these databases have identified more than 6,300 victims and nearly 3,200 offenders globally since 2009 (INTERPOL, 2015). Youth-produced sexual content is on the rise, and the IWF conducted a study in 2015 collecting data over three months and observed that 85.9% of youth-produced sexual content used laptop webcams, and 17.5% of the material depicted individuals under 15 years old (IWF, 2015a). The number of child abuse images had been estimated to be around one million, with as many as 50,000 new images going into circulation per year (ICMEC and Carr, 2017).

1.1.2 <u>**Using technology to prevent trafficking**</u>

In order to combat the vast amount of child sexual abuse images, companies and organisations have used a variety of methods to detect illicit material on their systems (ICMEC, 2013). The Internet Watch Foundation (IWF) reported a 137% increase in the identification and assisted removal of web pages containing child sexual abuse material in 2015 compared to 2014, where less than 0.3% of child abuse content was being hosted within the UK and 95% of these web pages were removed within a day (IWF, 2015b). In 2017, less than 1% of CSAM was hosted in the UK (IWF, 2018). NCMEC (2015a) received 4,403,657 CyberTipline reports, which was a 298% increase compared to 2014, and 99% of those reports were related to indecent images involving children.

With the emergence of cloud computing, and an increased use of the hidden internet to disguise identity and encrypt the sharing of child sexual abuse images; hashing technology was able to identify and block these images from being shared by using digital fingerprints, also known as hash values (GOV.UK, 2014; ICMEC and Carr, 2017). Known child sexual abuse images identified by the IWF can help to prevent sharing on companies such as Facebook, Microsoft, Google, Twitter and Yahoo (GOV.UK, 2014). Companies and organisations such as Interpol have used Microsoft’s PhotoDNA to calculate the hash of images for comparison to the ICSE database, and the use of hash technology can eliminate the duplication of images within the database and help speed up the identification process (ICMEC, 2013; INTERPOL, 2015).

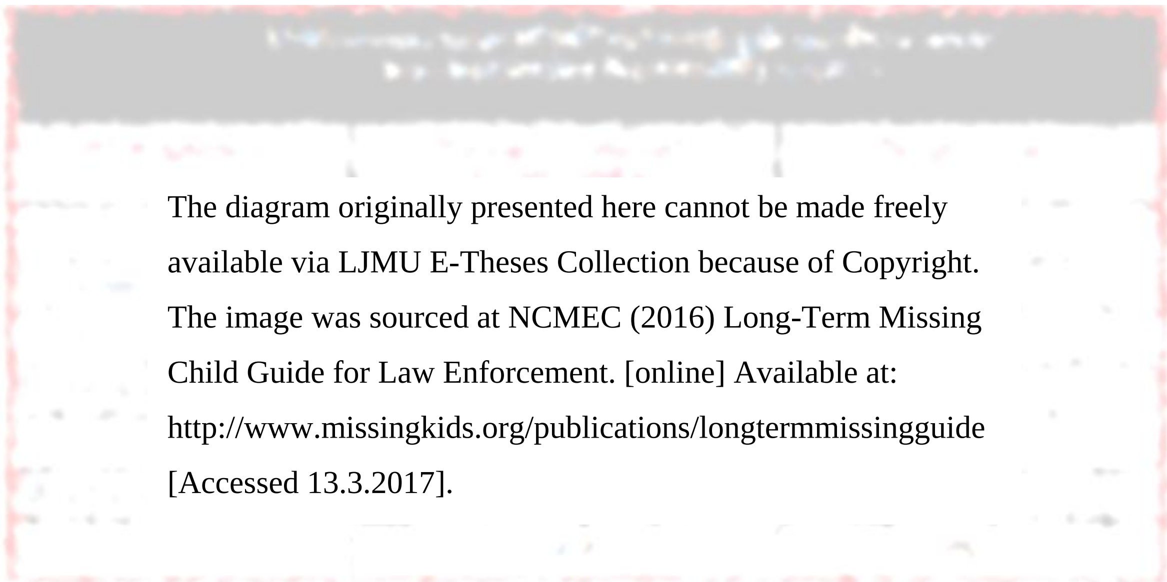
The sexual exploitation of children can be reduced by limiting the source of material. Different law enforcement organisations have formed an alliance to protect children from online sexual exploitation. The Virtual Global Taskforce (VGT) was formed by 14 organisations across the world (VGT, 2011), and since 2003, they have helped identify sex offenders with projects such as operation PIN. Projects like this aimed to capture information relating to paedophiles by setting up a fake website claiming to contain CSAM (Wei, 2012). Law enforcement agencies have used programs such as Fairplay and RoundUp to identify IP addresses in peer to peer distribution of child sexual abuse image files, and these programs helped identify 20 million addresses between 2006 and 2009 (U.S. Department of Justice, 2010).

CSAM can be filtered by internet blocking or through notice and takedown where members of the public report sites containing CSAM through hotlines established in different countries (ICMEC and Carr, 2017; Wei, 2012). With 1.2 million reports received in 2013, international collaboration is on the rise, with networks such as INHOPE having 51 hotlines across 45 countries. By developing a secure software to collect, exchange and categorise reports on CSAM, these networks work together to remove illegal content and prevent distribution and circulation of such material to protect child victims (INHOPE, 2018, 2014). However, different standards across countries can result in the inefficient takedown of CSAM. For example, the definition of a child differs and the definition of CSAM can also vary between different countries (Wei, 2012). In addition, many places do not have these systems in place, and CSAM can still be exchanged using other means, such as peer to peer file sharing, email and free hosting sites (Wei, 2012).

1.1.3 Recovery rate

NCMEC in the USA has assisted in the identification of 2,589 children related to indecent images, and the centre has recovered more than 205,550 children since 1984 with a recovery rate of 97%, a 35% increase when compared to 1990 (NCMEC, 2015b, 2015a). China started a project in 2007, called 'Baobeihuijia', to reunite missing children and their families (www.Baobeihuijia.com). The project advised the general public to take photographs of lost or street children through a mobile phone application, and these photographs were compared to the database of missing children using facial recognition software (Yao, 2014). Since 2007, this project has helped 1,406 missing children to reunite with their families (Baobeihuiji, 2016). Similarly, in India, the TrackCHILD facial recognition system has helped the Ministry of Woman and Child Development (2013) in the identification of 2,930 children from 45,000 photos (John, 2017; Kovner, 2018; Marchildon, 2018; NDTV, 2018).

In 2015, NCMEC distributed 20,230 photos of missing US children (NCMEC, 2015a). With the aid of technology, it is becoming more common to find long-term missing children (NCMEC, 2016). Figure 4 shows the number of recoveries between 2011 and 2015. Although the 2015 report did not specify the statistics on the methods leading to the identification, NCMEC readily used age progression and sophisticated forensic technology to search for missing children.



The diagram originally presented here cannot be made freely available via LJMU E-Theses Collection because of Copyright. The image was sourced at NCMEC (2016) Long-Term Missing Child Guide for Law Enforcement. [online] Available at: http://www.missingkids.org/publications/longtermmissingguide [Accessed 13.3.2017].

Figure 4: National Center for Missing & Exploited Children (NCMEC, 2016) long-term US missing children recovery figures between 2011 and 2015

The world's internet usage reached 3 billion (42.3%) in 2014 (Miniwatts Marketing Group, 2015), and the platform of social media usage has grown exponentially over the past few years. Facebook remained one of the most used sites with the number of active users reaching 1.3 billion, QZone from China was in second place with 0.6 billion, and others, such as Google+, LinkedIn, Instagram, Twitter, Tumblr etc. (Statistic Brain Research Institute, 2015a;b) were not far behind. In a recent social experiment, a photographer took images of strangers and using just a facial image, he was able to find out information about the stranger on social media using a website called 'Find Face' (McGoogan, 2016). It is therefore possible that victims of child trafficking from illegal adoption, sexual exploitation, forced labour and other forms of exploitation could appear on social media. Facebook holds more than 250 billion photographs and with more than 4.75 billion items being shared each day (Facebook et al., 2013), these data could help to find missing children. However, the identification rate of children from OCSE is low, especially when the location of the child and the offence are often unknown (ICMEC and Carr, 2017).

1.2 Literature Review and Application

Face recognition systems can be classified as controlled and unconstrained. In practice, *controlled face recognition* is achieving near perfect accuracy; it is often used in biometric systems for identity verification, where recognition is in a controlled environment with standardised illumination pose and facial expression (Hassan et al., 2015). *Unconstrained face recognition* is more challenging with variations in illumination, pose, facial expression and also the quality of the images (Hassan et al., 2015). Although not as accurate as controlled face recognition, reported accuracy is advancing with recent research involving deep learning. However, unconstrained face recognition continues to challenge this field of research (Hassan et al., 2015).

1.2.1 Facial recognition systems (FRS)

All facial recognition systems/algorithms (FRS) are developed using a database of faces, and these datasets vary in the number of photographs, the number of individuals and the conditions. There are many publically available datasets, but they are relatively small in comparison to the training datasets used by big companies, such as Google who have access to at least 100-200 million photographs of 8M individuals (Schroff et al., 2015) or Facebook who have access to at least 4.4 million photographs of 4K individuals (Taigman et al., 2014). Miller and colleagues (2015) tested four different types of algorithms along with human performance (Figure 5), and they found that by adding a larger dataset (maximum 1M) FaceNet (developed by Google) was the most robust achieving 75% identification rate even with 1M distractors, whereas other algorithms dropped by 70%. This drop in recognition was somewhat surprising since most reported a high recognition rate when tested on the 'Labeled Faces in the Wild' (LFW) dataset (Table 1). This result indicates that the size of the training dataset is crucial for the algorithm to learn and distinguish the difference between similar faces.

The diagram originally presented here cannot be made freely available via LJMU E-Theses Collection because of Copyright. The image was sourced at Miller, D., Kemelmacher-Shlizerman, I. and Seitz, S.M. (2015) MegaFace: A Million Faces for Recognition at Scale. arXiv:1505.02108.

Figure 5: Face Recognition performance with added distractors (Miller et al., 2015)

'Labelled Faces in the Wild' (LFW) is a public dataset containing more than 13K unconstrained facial images collected from the internet. This dataset had been widely used to test the performance of the FRS. Table 1 shows some recent recognition systems tested on the LFW with a recognition rate close to the human performance, numbers 1, 3 and 4 were tested on the MegaFace database (100M faces from Flickr) by Miller et al. (2015) from the database YFCC100M (Thomee et al., 2016):

Table 1: Published recognition rates for Face Recognition systems tested on the 'Labelled Faces in the Wild' dataset

	Name or method of system	Institution or company	Recognition rate on LFW	Citation
1	FaceNet	Google	99.63%	(Schroff et al., 2015)
2	GaussianFace	Chinese University of Hong Kong	98.52%	(Lu & Tang, 2014)
3	Joint Bayes	CASIA*	97.73%	(Yi et al., 2014)
4	Human Performance		97.53%	(Kumar et al., 2009)
5	DeepFace	Facebook	97.5%	(Taigman et al., 2014)

** Center for Biometrics and Security Research & National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences (CASIA)*

Recent research in facial recognition has focused on building a large database of faces (Miller et al., 2015; Yi et al., 2014), as researchers believe that the available dataset for training could be more important than the algorithm (Yi et al., 2014). However, Grother and Ngan (2014) suggest otherwise, and state that recognition accuracy was dependent on the algorithm, specifically the developer. Mahmood et al. (2016) compared three different baseline algorithms against pose variation and low-image resolution and suggest that some algorithms were more robust against these different factors than others. For example, PCA is optimal with pose variation and AdaBoost was optimal at the identification from low-resolution images. This suggests the success of an algorithm is not solely based on the size of the training base, but also on the engineering of the algorithm.

FRS such as FaceNet were trained on very large databases, and they outperform others when tested on the MegaFace Dataset (Miller et al., 2015). The larger the database, the higher the probability of having similar faces, and this will result in an increase of false positive and false negative identifications (Grother and Ngan, 2014). A big training dataset is important to the success of the algorithm (Parkhi et al., 2015), and the bigger the data, the more sensitively the algorithm can be trained to distinguish similar faces. Algorithms developed by Google (FaceNet) and Facebook (DeepFace) both involve Deep Convolutional Neural Network (DCNN), a form of deep learning (Rawat and Wang, 2017; Schroff et al., 2015; Taigman et al., 2014).

Deep learning is a powerful tool for modern-day machine learning, as it is able to train neural networks to learn and recognise patterns when adequate examples are provided (Hassan et al., 2015). DCNN, in particular, became the leading method for different analysis of imagery (Phillips et al., 2018; Ranjan et al., 2017; Rawat and Wang, 2017). As one of the frontiers in neural networks, it is arguable that the algorithm Google developed is able to perform much better in comparison to others, but the fact that the developer trained the algorithm using a large database could be a significant factor contributing to its success.

The National Institute of Standards and Technology (NIST) have developed standardised tests to assess the performance of commercial facial recognition software on a database of 1.6 million mugshots (Grother and Ngan, 2014; Grother et al., 2010). When using good quality mugshots, a commercial algorithm developed by NEC performed the best with 4.1% of the identifications failing to be in a rank-1 position (top one) and 2.6% failing to be in rank-5 (top five) (Grother and Ngan, 2014). Algorithms are able to recognise faces under controlled conditions with high accuracy, but recognition becomes much more challenging when unstandardised (unposed/unconstrained) faces are utilised (Bourlai, 2016). Missed identification at rank-1 increased to 20-60% when poor quality webcam images were used (Grother and Ngan, 2014). This suggests that the image quality is also a determinant factor in the success of an algorithm. In reality, indecent images of children will not be of high quality.

1.2.2 Facial recognition in children

How good are FRS at recognising the same face years apart? Ling et al. (2010) designed a face verification algorithm and tested faces across different ages for children and adults. Their study found that verification was much harder for children in comparison to adult faces and it was extremely difficult to verify the identity of children between 0-8 years of age. This is unsurprising, as the algorithm considers the face to be a universal, distinctive, permanent and collectable biometric (Jain et al., 2004b). However, as children's faces change rapidly over short periods of time, facial recognition in children cannot be classed as a reliable biometric method, as facial characteristics are invariant. Some researchers consider a child's face as a soft biometric (Matthews et al., 2018) and it was defined as having *"characteristics that provide some information about the individual but lacks the distinctiveness and permanence to identify an individual uniquely and reliably"* (Jain et al., 2004a). Humans often identify each other with soft biometric traits, for example, height, weight, gender, eye colour, ethnicity etc. (Jain et al., 2004a; Reid et al., 2013; Reid and Nixon, 2011). Ferguson (2015) suggested that the manual facial comparison of juvenile faces is error prone. If we were to consider a child's face as a soft biometric, we would need to consider the human ability to recognise children's faces even when they are years apart. How good are facial recognition systems in the identification of children across time? Perhaps identification with a focus on stable features and facial markings such as moles should be evaluated further (Caplova et al., 2017).

The National Institute of Standards and Technology (NIST) reported that the false negative and false positive rates for juvenile FRS were much higher than for adults. They found a progressive trend in the decrease of false identification with increasing age and concluded that it was difficult to discriminate younger children, with a high false positive rate across all algorithms (Grother and Ngan, 2014). It is not known whether the training dataset for these algorithms contained images of young subjects as the companies do not publish this information. The report suggested, "*Younger children are more difficult to discriminate"* and this could suggest that younger children look similar to each other. It is important to know if the algorithm would perform better if it was trained to distinguish younger individuals and it would be interesting to see if the algorithm can perform any better when developed on a database with younger subjects.

1.2.3 Verification and Identification

Law enforcement has attempted to test and quantify the capability of FRS to detect and recognise the unconstrained faces of children. The Child Exploitation Image Analytics (CHEX-IA) was an imagery evaluation from the NIST; this CHEXIA-FACE test recruited universities and commercial entities to participate their FRS in four categories: identity (1:1) verification, large-scale (1:n) identification, face detection, and clustering of images (Grother and Ngan, 2015).

In the literature and in biometric technologies, most systems have demonstrated 1:1 verification by comparing one source images to one target, and 1:n identification by comparing one source images against a collected database of images. This golden rule is useful when dealing with simple pattern recognition such as fingerprint, iris pattern and perhaps even the standardised frontal view of an adult face. In situations such as for images related to child exploitation, it is unlikely that the source images and the target images will be taken in a standardised environment. In addition, facial changes due to growth will provide even more challenging situations when one source image is utilised for recognition. However, the meaning of '1' in 1:n verification from the CHEXIA-FACE can sometimes contain multiple images of the same individual in a combined template (Grother and Ngan, 2015).

Digital photography has become widespread over the past decade, and images are taken with ease and with increased frequency. For images taken 'in the wild', the differences between individuals could be diminished with the wide variations in pose, lighting etc., and identification within a large dataset for these images will be even more challenging (Stone et al., 2010). With the increase in memory storage and the continuous improvements in the quality of digital photographs, photo management applications have increased in popularity (Cui et al., 2007). These applications, such as Google Picasa, Flickr and Facebook, have facilitated the development of features such as automated face detection, face tagging and clustering. These management features could be useful when dealing with facial images taken in unstandardised conditions ('in the wild'), such as images related to child exploitation. When more faces are tagged for each individual, it will be statistically more likely for their face to be identified in a large database of images. In situations with unconstrained facial images, using multiple 'source images' for facial recognition was shown to be beneficial (Mu et al., 2014; Schroff et al., 2015); therefore, using more than one image of the child could potentially improve recognition across different ages. Using more than one age progression depiction could also be beneficial to recognition (Lanitis and Tsapatsoulis, 2016).

This study demonstrates the use of a commercial photo management application (Google Picasa) in identity verification across the different ages of the same individual; the use of multiple source images will be compared to a single source, and the limit of face recognition in relation to facial change will also be explored. How large of an age difference is necessary before the FRS fails to recognise the child as the same individual?

1.3 Age progression

To model and predict the possible changes to an ageing face, age progression methods change the shape, colour and texture of a facial image while retaining the identity of the individual (Hunter et al., 2012). The areas of change are different for adults and children, and during juvenile growth, the skull and associated cartilages change in size and proportion to accommodate the growth and development of the internal organs (e.g. the brain, airway, dentition etc.) and increased body size. However, with the skull shape remaining relatively stable in adulthood, the changes in face shape relate to the continuous growth of cartilage (i.e. the nose and ears) and soft tissue changes, such as the development of wrinkles and skin sagging. Therefore, age progression is often separated into juvenile and adult (Mullins, 2012), with many freely available adult ageing applications or programs such as HourFace (MotionPortrait, Inc., 2015) or in20years (Luxand, Inc., 2015). This study focuses on age progression for juvenile faces. Age progression is challenging for individuals younger than 3 years of age, as facial characteristics are underdeveloped at this stage in the growth pattern (Mullins, 2012). Age progression is more accurate with images of older children and accuracy is also affected by the quality of the reference photographs (Mullins, 2012). Current research techniques include manual or machine-based digital image processing and sometimes drawings by artists (Mullins, 2012). NCMEC in the USA updates the age-progression image every 2 years before age 18 years, and every 5 years after age 18 years. These images are used to generate further investigative leads (NCMEC, 2016).

1.3.1 Machine-based age progression

Previous literature has described machine-based age progression methods as automated or computerised methods. The level of automation of age progression is still in its infancy and requires a high level of human influence.

Different research groups have developed methods to automate age progression using various algorithms, and most methods are based on averaged anthropometric growth patterns (Lampinen et al., 2010). A few approaches are discussed below:

1.3.1.1 2-Dimensional age progression models

Ramanathan and Chellappa (2006) and Wu and Chellappa (2012) developed craniofacial growth models using sets of linear equations based on craniofacial anthropometry in relative growth parameters (Figure 6). This is similar to cardioidal strain based methods where the head shape changes related to bone growth (Hunter et al., 2012). This mathematical model was developed to verify age-separated images of individuals under 18 years of age. Recognition becomes even more challenging in children under 15 years of age due to the rapid growth of the face, especially with features along the outer contour (Ramanathan and Chellappa, 2006).

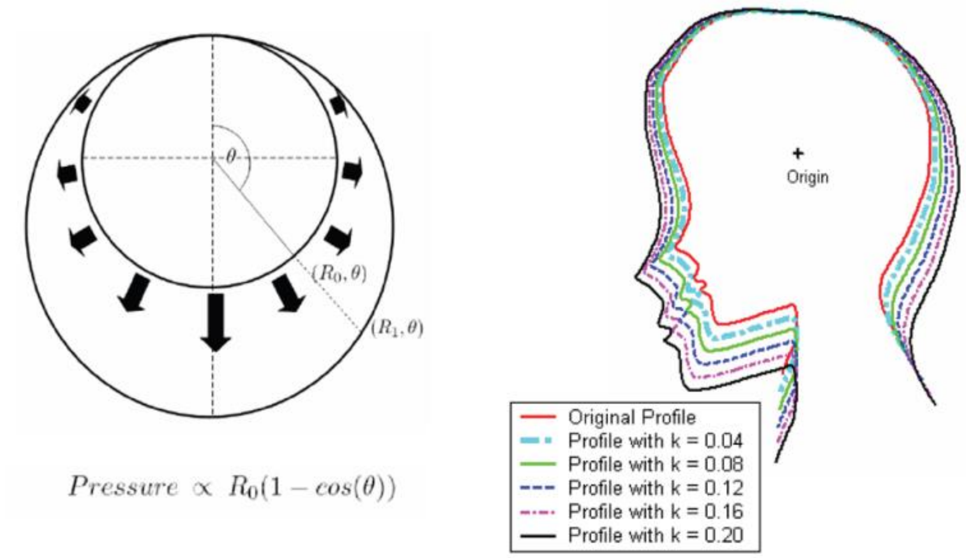


Figure 6: Cardioidal strain based generative method for facial growth in children (Ramanathan et al., 2009) © 2006 IEEE.

The model progressed and aged a single 2D image of the face based on male and female anthropometric proportion indices (Ramanathan and Chellappa, 2006). The distance between the eyes remains relatively stable after infancy; therefore, the midpoint of this distance was used as a reference point to set the coordinates for alignment. The region of the eyes also remains relatively stable during the growth process, and the contour of the face, shape of the nose and mouth change more in comparison. The results from age-separated images suggested that facial recognition was extremely difficult for children under age 8 years, but by using the growth model, prediction of new facial shapes in teenagers showed an increased facial recognition performance (Wu and Chellappa, 2012). However, this growth model relied on the known age of the source image, and this information could be inaccurately reported by the donor. In addition, the model developed may only be Caucasian-specific and it also lacked age-related changes such as skin texture, facial hair and facial fat distribution (Ramanathan and Chellappa, 2006). Finally, although cardioidal strain based methods may work well for large proportional shape changes of the face, it was not able to account for colour and textural changes to the face (Hunter et al., 2012).

Kemelmacher-Shlizerman et al. (2014) developed the illumination-aware age progression technique using subspace-to-subspace alignment, and their workflow was able to produce a series of age-progression images from a single photograph of a child with 4 steps (Figure 7). First, to account for pose difference, the original image was corrected to the frontal pose, secondly, the texture was relit to match the target age. The process followed by applying the flow difference between the source and the target age, and finally, the aspect ratio for the difference in head shape due to ageing was adjusted. To demonstrate the ability to match the illumination of another image, the authors also used the actual images of the target age as the relighting reference. The study suggested that this ground-truth-blended comparison performed better for ageing children when compared to other methods. As a result of the blend, the outer features between the progression and the target image was identical (i.e. hairstyle and clothing) which could create a bias when testing recognition. This method was developed on 40K 'cross-sectional' images across different ages. This method also accounted for shape, colour and texture changes of the face whilst retaining the original images from the individual; this could potentially increase the accuracy in modelling age-related changes compared to previous methods described above.

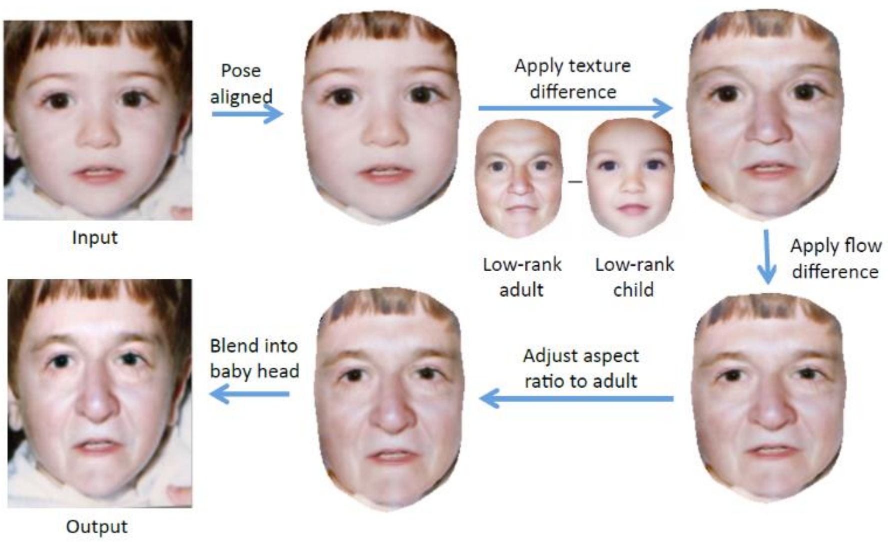


Figure 7: Steps of illumination-aware age progression (Kemelmacher-Shlizerman et al., 2014) © 2014 IEEE.

Machine-based age progression studies have focused on the generation of realistic textures, rather than relying on an input image to match the illumination (Kemelmacher-Shlizerman et al., 2014). Bukar et al. (2017) proposed a framework using a hybrid technique and unlike other statistical models, this technique was able to create depictions with finer facial details. High quality coloured images with varying facial expression and head poses were collected into nine age groups for texture enhancement implementation (Figure 8), and these were used to generate patch libraries for each age group. To generate a texture with fine details, the patches considered and overlapped small segments of the faces. This method eliminated illumination differences; whereas any gradient difference may remain using other illumination aware methods, such as the method proposed by Kemelmacher-Shlizerman et al. (2014).

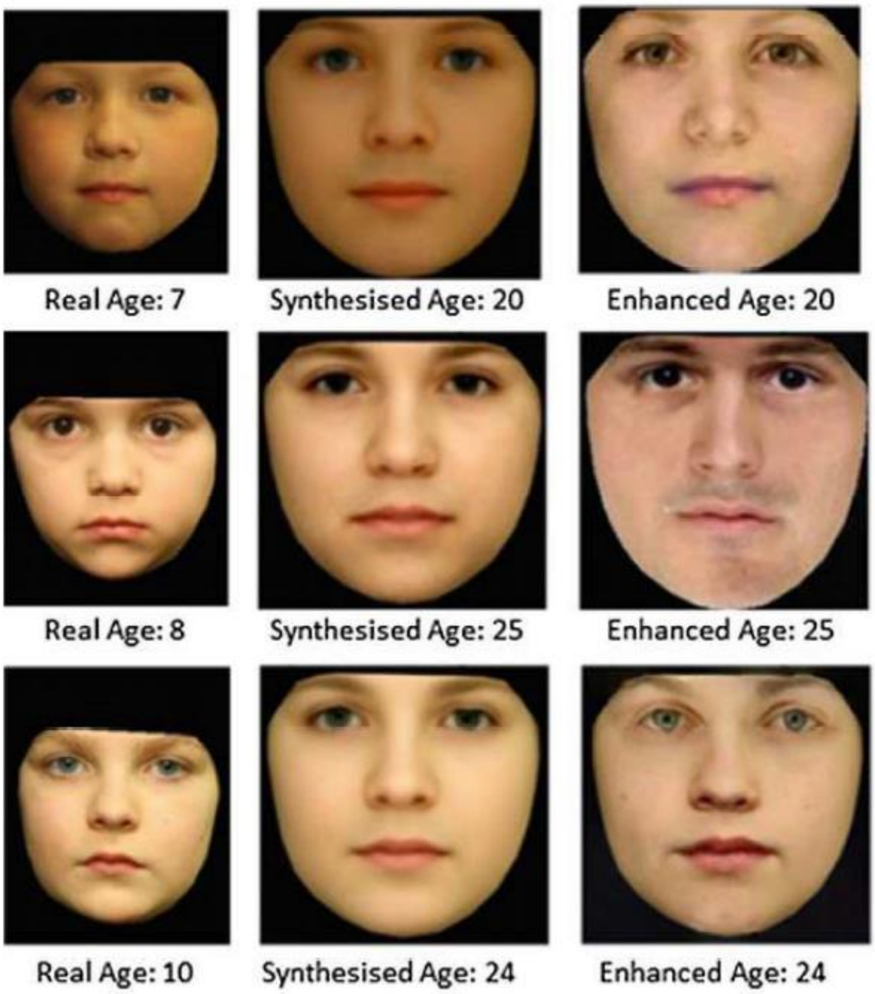


Figure 8: Texture enhanced age progression method (Bukar et al., 2017)

1.3.1.2 2D and 3D age progression models

To overcome the challenges of the unconstrained head pose between the original and the veridical image, some studies have developed machine-based age progression systems using morphable face models to match the difference in head pose between the two different unconstrained images (Scherbaum et al., 2007; Shen et al., 2014).

Scherbaum et al. (2007) developed a non-linear ageing curve for a machine-based age progression model based on 393 individuals between 8 and 30 years old (Figure 9). Faces of 238 teenagers between 8 to 16 years old were scanned, and these scans were used to develop the model. In combination with the high-quality digital images of the subjects, the texture was extracted to produce high-quality texture maps onto the model with the ability to match the illumination to the 'Ground-truth'; different hairstyles could also be applied.

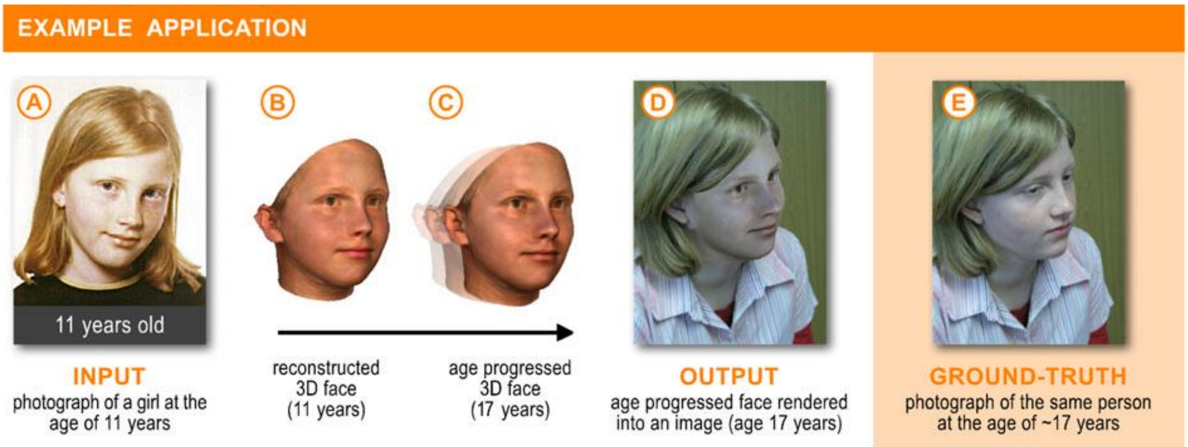
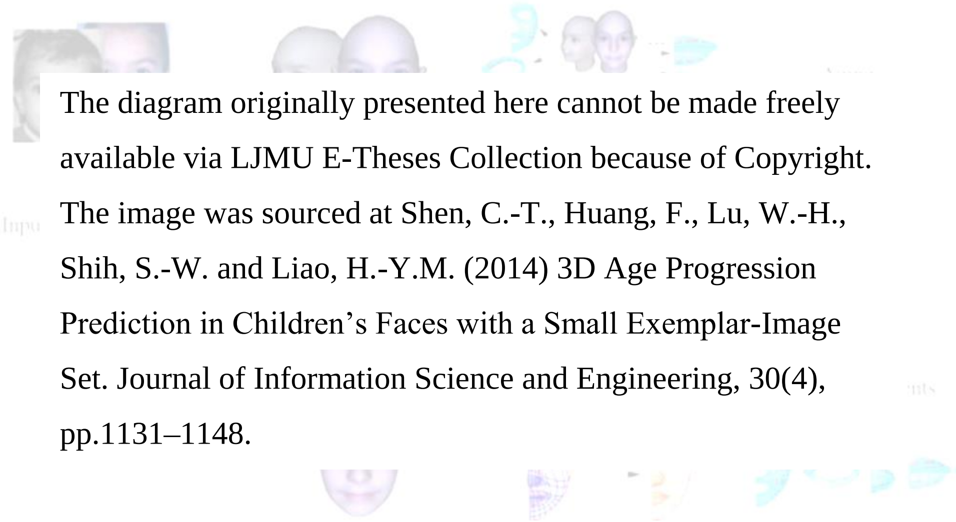


Figure 9: 2D/3D age progression method (Scherbaum et al., 2007)

Similarly, Shen et al. (2014) developed a machine-based age progression system using 3D face models. First, a 2D image of the child was converted into a 3D face. Then each facial component (face shape, eyes, nose, lips etc.) was extracted, and the growth curve of other children with similar faces was applied to each component individually (Figure 10). One of the biggest limitations noted by the author, was that only the FG-NET was used to train the algorithm to establish a growth model. This database contains less than 100 individuals across a wide variety of ages from age 0-69, which is not sufficient for a reliable age estimation. Although the design of this algorithm could be good for small training datasets, further testing and training are required.



The diagram originally presented here cannot be made freely available via LJMU E-Theses Collection because of Copyright. The image was sourced at Shen, C.-T., Huang, F., Lu, W.-H., Shih, S.-W. and Liao, H.-Y.M. (2014) 3D Age Progression Prediction in Children's Faces with a Small Exemplar-Image Set. Journal of Information Science and Engineering, 30(4), pp.1131–1148.

Figure 10: 2D/3D age progression method (Shen et al., 2014)

1.3.1.3 3-Dimensional age progression models

Koudelová et al. (2015) modelled age progression specifically in children between 12-15 years old (Figure 11) and developed a prediction model using geometric morphometric (GMM) based on 45 Caucasian 3D faces (23M; 22F). This was a longitudinal study where each individual had their face scanned at 4 consecutive years between 12-15 years old. The facial form showed a significant difference between the age groups for each sex by using principal component analysis (PCA), and the changes for boys were more prominent than the girls. The authors reported a mean error of 1.92mm in girls and 1.86mm in boys (Figure 11).

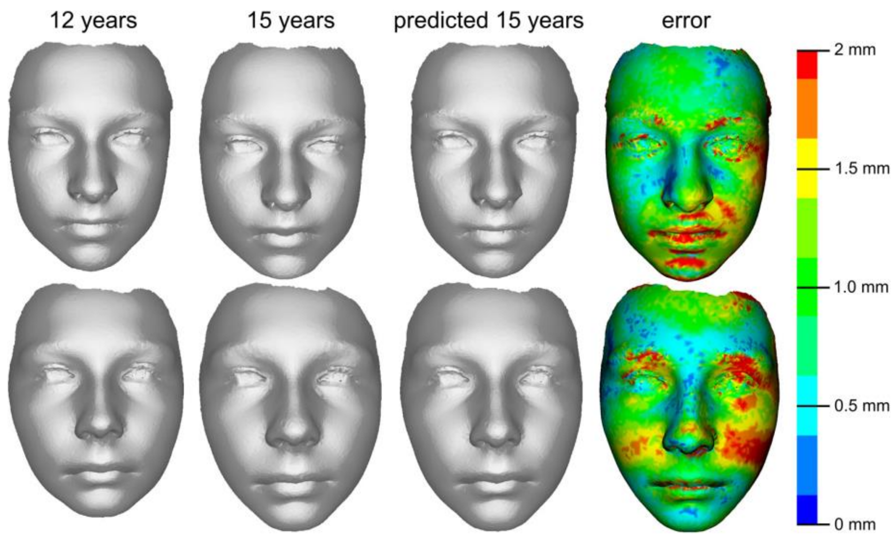


Figure 11: 3D modelled age progression method (Koudelová et al., 2015)

Matthews et al. (2018) developed a framework for machine-based age estimation and age progression. The algorithm was trained using cross-sectional 3D photographs (360 degrees) from individuals between 0 and 18 years old (452M; 422F), and the model was validated using a longitudinal subset of 50 subjects (24M;26F) who had been photographed at two different ages, with an interval between 3.61 - 6.40 years. The authors reported an average of 85.07% accuracy of the face, and 74.80% of the head within three millimetres (Figure 12).

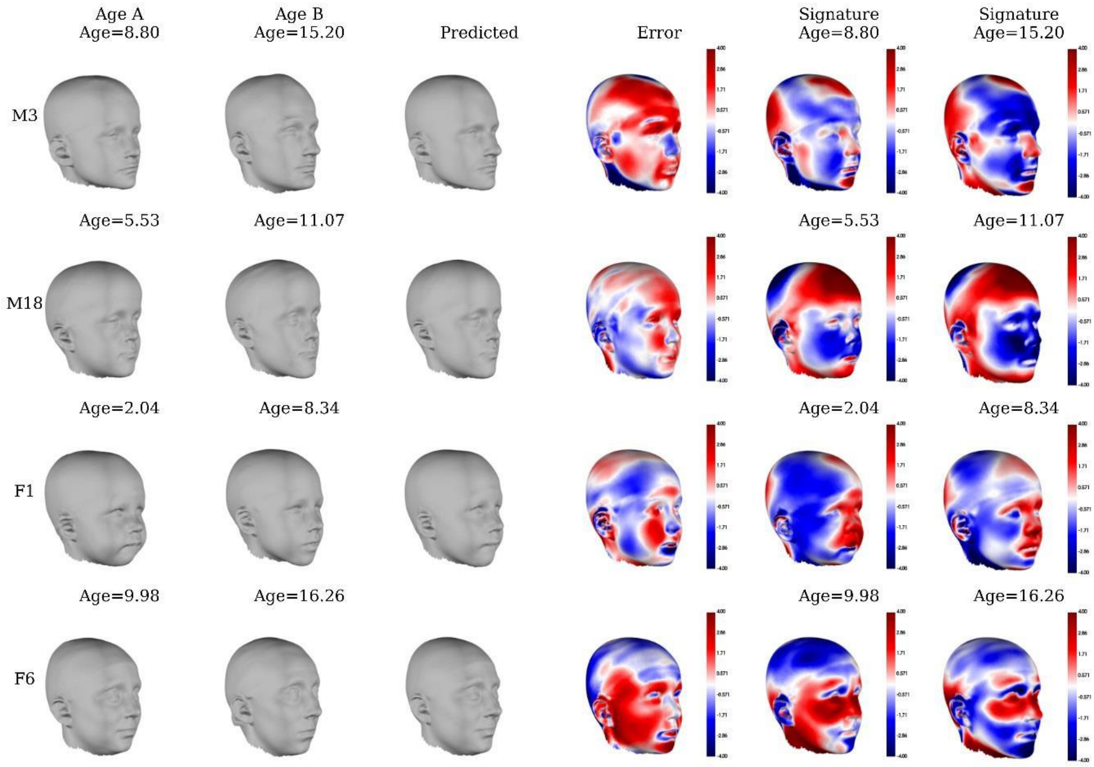


Figure 12: Synthetic facial growth maps for 8-15 years (Matthews et al., 2018)

** The scale of the colour deviation map: +- 4 millimetres

1.3.1.4 Comparison between 2D and 3D methods

In comparison to 2D studies using photographs (Kemelmacher-Shlizerman et al., 2014; Ramanathan and Chellappa, 2006), the 2D/3D (Scherbaum et al., 2007; Shen et al., 2014) and the 3D methods (Koudelová et al., 2015; Matthews et al., 2018) could be more beneficial. 2D photographs can often introduce perspective and projection error, especially when the training database is not standardised. Since 3D imaging is likely to increase in popularity (Matthews et al., 2018), 3D methods of age progression could be useful in populations where subjects have recorded 3D imaging before they went missing. A database with longitudinal 3D faces is able to hold more information, regarding the shape, true size and growth specific to each individual within the sample. Cross-sectional studies do not reveal the true growth for each individual. However, the collection of a longitudinal 3D database, as shown in Koudelová et al. (2015), is often time-consuming with a limited variation within the sample.

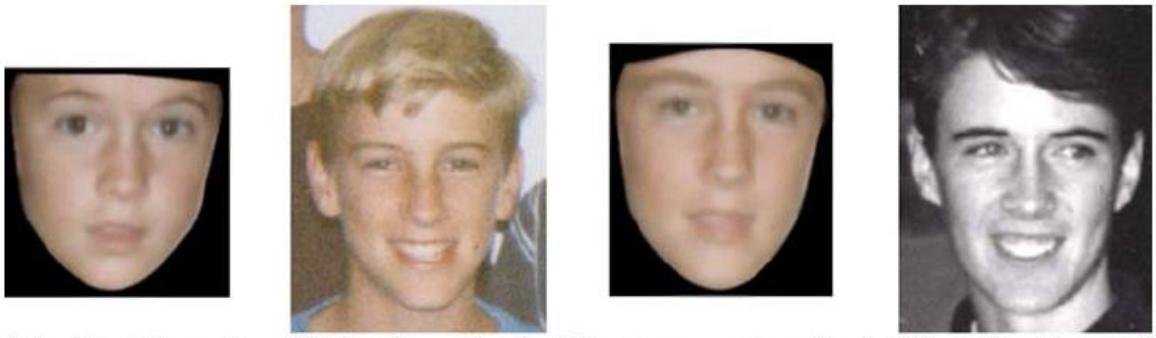
The age progression model from Koudelová et al. (2015) seems to be more age accurate in comparison to the Matthews et al. (2018) model. Apart from the difference in using longitudinal or cross-sectional data, a difference in age range and the area of interest could also be a major contributing factor in this difference. Matthews et al. (2018) tested larger age intervals at different age groups, whereas the Koudelová et al. sample was more controlled with a 3-year age interval of the same subjects. Koudelová et al. (2015) restricted the area of interest to the face only, whereas Matthews et al. (2018) used a whole head model. For the purpose of forensic age progression where the face is the most identifiable feature, focusing only on the face could be more beneficial. Matthews et al. (2018) noted that the shape of the overall head was less accurate in comparison to the face region.

Both texture and shape are important factors in facial recognition (O'Toole et al., 1999). Faces without details and colour, such as the 3D models produced by Koudelová et al. (2015) and Matthews et al. (2018) may achieve a lower recognition rate (Bruce et al., 2013). Texture can be applied to 3D models, but the 'wrong' texture can lead to incorrect recognition (Claes et al., 2010a). The 2D render of the 3D face model will also differ to a photographic image. Some studies have addressed this issue by matching the illumination of the image to the target image (Kemelmacher-Shlizerman et al., 2014; Scherbaum et al., 2007). If these age progressions perform better for a machine-based facial recognition system, this method could be a beneficial investigative tool if the model is able to cross-match to the illumination of all possible images within the database. This could be computationally expensive depending on the size of the database, therefore it will be interesting to see if the methods proposed by previous studies (Bukar and Ugail, 2017) using detailed texture could lead to better FR performance.

1.3.1.5 Genetic influence

Craniofacial development is related to genetics and environmental factors with certain facial parameters being more genetically controlled than others (Cakan et al., 2012). Forensic artists often use information from images of family members for the creation of the age progression image (Erickson et al., 2016; Lampinen et al., 2015; Taylor, 2000). The methodology of the machine-based studies can be disadvantageous when hereditary information is not considered.

Gibson et al. (2009) proposed a computer-assisted age progression algorithm by using a combination of averaged growth models and genetic information from reference images of relatives. The reference image of the relative was visually assessed for facial similarities to the subject (Figure 13). This method was able to bias the age progression to be more like the relative than the averaged mode, which could be an important step in creating a better depiction.



(a) subject B age 7 (b) brother of subject B age 14 (c) age progression of subject B to 17 (d) veridical image of subject at 17

Figure 13: Computer-assisted age progression (Gibson et al., 2009)

Reprinted by permission from Springer Nature Customer Service Centre GmbH © 2009

Different approaches to age progression have been attempted, but “currently there is no automatic age progression software that can guarantee any degree of accuracy” (NCMEC, 2016). Although the literature has explored the human recognition rate using age progression images (see section below), but no published literature has reported the testing of automated recognition rates using the age progression and veridical (Target) image.

1.3.2 Manual age progression

A specialised forensic artist often creates manual age progressions by sketching or by utilising photo-editing software (e.g. Photoshop). The age progression technique can vary between different practitioners (Figure 14) (Erickson et al., 2016) and some practitioners prefer to put more weight on quantifiable growth data, whilst others put more weight on the features of other family members (Taylor, 2000). To understand and produce a more accurate depiction, images of siblings and parents at the same age of the progression are often required to help artists to maintain a reliable likeness with biological resemblance (Lampinen et al., 2015). But when these images are not available, a more general reference will be used, such as images of other children of the same age (Mullins, 2012).

Techniques and tools used	Artist							
	1	2	3	4	5	6	7	8
Growth norm database		x				x		
Personal growth norm knowledge	x		x	x	x	x	x	x
Biological relative photos at target age and last known photo	x	x	x	x	x	x	x	x
Lifestyle information	x	x	x	x	x		x	x
Medical information	x	x	x	x	x	x	x	x
Computer algorithms						x		
Hand sketches	x			x				x
Photoshop (or similar editing software)	x	x	x	x	x	x	x	x
Other (describe)			x[a]					

Note:
[a] "Geographical information".

Figure 14: Manual age progression techniques (Erickson et al., 2016)

© 2017, with permission from Elsevier

The original images should be altered as little as possible to retain certain facial characteristics, by using reference material of other children, only small portions should be used to avoid resemblance from the templates (Mullins, 2012). Manual age progression methods are subjective, not standardised and also vary between forensic artists (Charman and Carol, 2012; Koudelová et al., 2015; Lampinen et al., 2015). Understanding the growth in children will be beneficial when changing the proportion of the head and face in an age progression (Farkas et al., 1994; Taylor, 2000).

The most important proportional change is the lower face growing in length and width and prominence (Taylor, 2000). Based on a cross-sectional Caucasian sample (n=2326), Farkas and Hreczko (1994) provided a set of growth-related linear measurements of the head, face, orbits, nose, lips and mouth. The authors provided: a comparison of measurements at age one; the total growth difference between ages 1 to 18 years; periods of rapid growth; and the maturation age in each individual measurement. These measurements could be useful to determine the parameter of change required for specific areas of the face. For example, the length of the head matures at around age 10 years for females. Information like this could provide a more 'guided' process of age progression. No matter what method is used, the practitioner must have a good knowledge of craniofacial growth and dental eruption patterns (Taylor, 2000).

1.3.2.1 Facial anthropometry

Farkas and Hreczko (1994) measured growth-related changes in North American Caucasian subjects across ages 1 to 18 years old (Cross-sectional). Numerous measurements of the head, face, orbits, nose, ears, lips and mouth were recorded from each year group, mostly between age 1 to age 18 years. The authors showed the difference in measurements between age 1 and 18 years as the total growth increments, and the period of accelerated growth in each region. In the majority of measurements, the authors showed that females had an earlier maturation rate in comparison to males. Of all the measurements between ages 1-18 years, most changes (growth over 20mm) lie within the head and face as listed in Table 2 and Figure 15 below:

Table 2: Growth changes (>20mm) from age 1-18 years

(Amended from Farkas and Hreczko 1994)

Linear Measurements	Total Growth between 1&18 years mean		Maturation age (Years)	
	mm	% **	Male	Female
Face: Mandibular arc (t-gn-t) *curve line	68.8	30.49	15	14
Head: Craniofacial Height (v-gn)	49.66	28.27	15	11
Face: Maxillary arc (t-sn-t) *curve line	49.6	22.16	14	12
Face: Width (zy-zy)	37.4	38.90	15	13
Face: Depth in Mandibular region (t-gn)	34.7	35.02	15	13
Face: Height (n-gn)	30.7	38.91	15	13
Face: Depth in Maxillary region (t-sn)	28.6	30.65	14	12
Head: Length (g-op)	24.5	14.91	14	10
Head: Width (eu-eu)	23.9	19.31	15	14
Face: upper face height (n-sto)	23.3	48.80	14	12
Nose : Height (n-sn)	20.9	69.55	15	12
Nose : Bridge length (n-prn)	20.5	76.64	15	13

***The total growth in percentage was expressed [Growth difference (mm)/ Mean value at age 1 years] %*

The diagram originally presented here cannot be made freely available via LJMU E-Theses Collection because of Copyright. The image was sourced at Farkas, L.G. and Hreczko, T. (1994) Age-related changes in selected linear and angular measurements of the craniofacial complex in healthy North American Caucasians, in: Farkas, L.G. (Ed.), Anthropometry of the Head and Face. New York: Raven Press, pp. 89–102.

Figure 15: Linear measurements with facial growth changes >20mm Amended from Farkas and Hreczko (1994)

Changes below 20mm from Farkas and Hreczko (1994) were mostly around the orbits and the mouth as listed in Table 3and Figure 16.

Table 3: Facial growth changes (<20mm) from age 1-18 years (Amended from Farkas and Hreczko 1994)

Linear Measurements	Total Growth between 1&18 years mean		Maturation age (Years)	
	mm	% **	Male	Female
Face: Width of the mandible (go-go)	18.7	24.80	13	12
Head: Height of the head (v-n)	18.5	19.14	13	13
Mouth: Width of the mouth (ch-ch)	17.5	51.40	14	14
Face: Height of the mandible (sto-gn)	16.0	50.55	15	12
Nose: Nasal ala length, left (ac-prn)	13.2	67.69	15	13
Orbits: Biocular width (ex-ex)	12.5	16.52	15	13
Nose: Nasal tip protrusion (sn-prn)	9.8	**96.55**	16	14
Nose: Width of the nose (al-al)	6.9	26.34	14	12
Orbits: Eye fissure length (ex-en)	5.3	20.66	15	13
Orbits: Intercanthal width (en-en)	5.2	19.19	11	8
Mouth: Height of the lower lip (sto-sl)	4.8	36.92	13	9
Mouth: Height of the upper lip (sn-sto)	3.9	23.15	11	5
Orbits: Eye fissure height (ps-pi)	1.2	12.57	11	14

***The total growth in percentage was expressed [Growth difference (mm)/ Mean value at age 1 years] %*

The diagram originally presented here cannot be made freely available via LJMU E-Theses Collection because of Copyright. The image was sourced at Farkas, L.G. and Hreczko, T. (1994) Age-related changes in selected linear and angular measurements of the craniofacial complex in healthy North American Caucasians, in: Farkas, L.G. (Ed.), Anthropometry of the Head and Face. New York: Raven Press, pp. 89–102.

Figure 16: Linear measurements with facial growth changes <20mm
Amended from Farkas and Hreczko (1994)

By separating the changes above and below 20mm, the practitioner can have a visual idea of the large changes relating to the facial growth. For example, sn-prn (nasal prominence) is a small measurement with a 9.8mm difference from age 1 to 18 years of age, but this change was nearly double the original measurement at age one (Table 3). Table 2 and Table 3 depicts the averaged measurements between male and female from Farkas (1994), with each measurement documented across the different age groups up to age 19-25 years old. This can be particularly useful in age progression, where measurements are taken from the photograph of the missing child (Farkas et al.,1994), and the known age is extrapolated according to the measurements from appendix A of Farkas (1994).

1.3.2.2 Iris ratio

Machado et al. (2017) analysed 10 facial measurements (Figure 17) from passport photographs of 1000 Brazilian subjects (n=200) age between 6-22 years. The authors compared nine different measurements of the face using the iris diameter as a fixed reference point. In comparison to interpupillary distance, the authors suggested that the diameter of the iris was the most stable measurement and could be a better reference for facial analysis. This can be particularly useful, as current age progressions are mostly digital, using tools such as Adobe Photoshop where the true measurement/scale is unknown. Farkas et al. (1994) used the endocathion distance (en-en) and the height of the upper lip (sn-sto) as a reference point for scaling the photograph to life-size in order to carry out measurements. Iris diameter could be a more stable reference point for standards in comparison to the method proposed in Farkas et al. (1994).

Nine out of ten anthropometry measurements from Machado et al. (2017) could be found in appendix A of Farkas (1994). With digital measurements taken in pixels, this makes comparison with anthropometric studies difficult. Anthropometry, such as Farkas (1994), are recorded as life-size measurements, and there will be differences when these measurements are translated to photographs, where the images are often affected by focal distance, distortion, head pose, facial expression, accessories such as glasses etc.

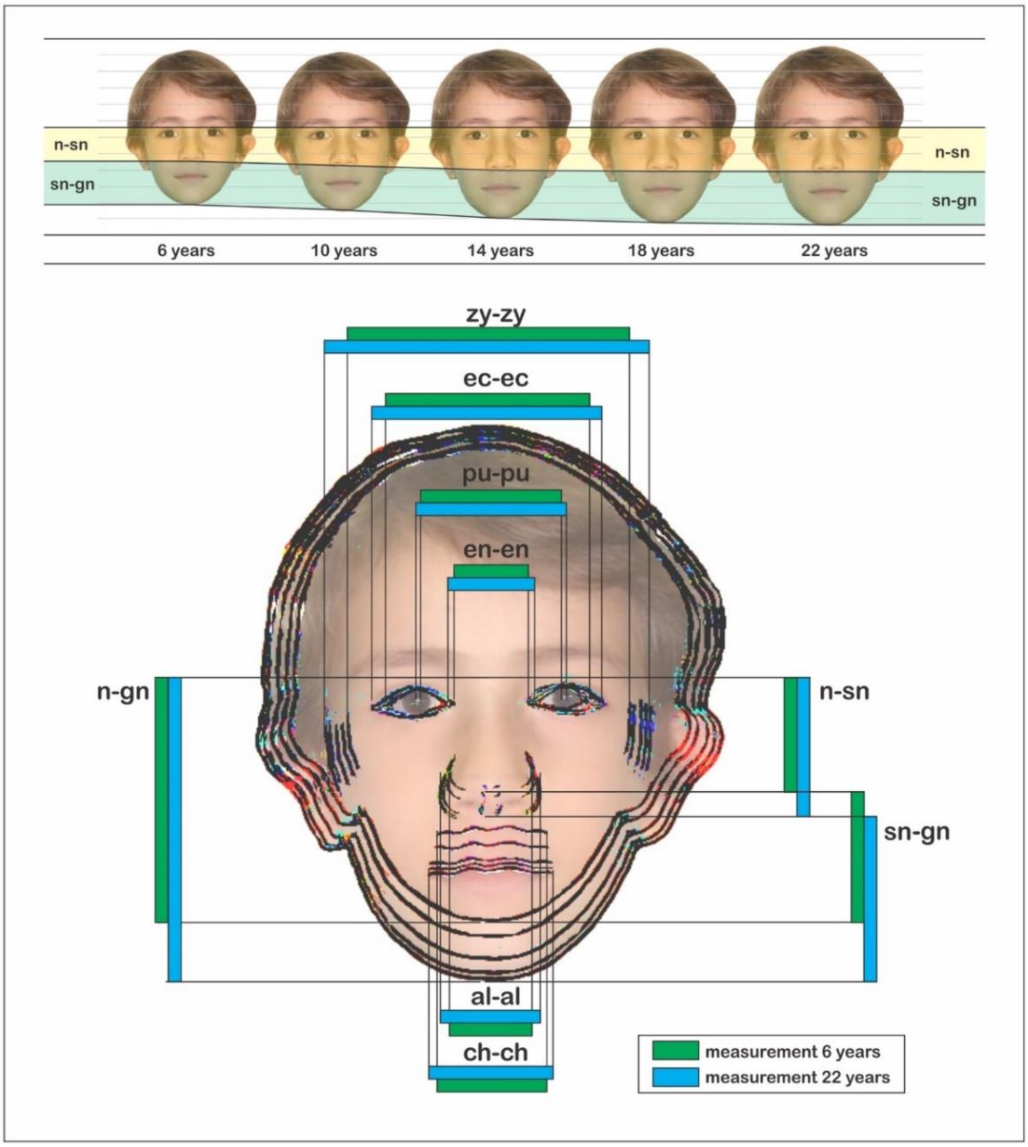


Figure 17: Craniofacial measurements (© Machado et al., 2017)

1.3.2.3 Comparison of anthropometric methods

Anthropometric measurements from Farkas and Hreczko (1994) could be a useful tool for age progression. However, these measurements are not directly translatable to photographic images unless we know the life size of the image (Farkas et al., 1994). The iris diameter in infants to 8 years of age showed a mean difference of 0.318 ± 0.10 S.E. mm (Ronneburger et al., 2006). Machado et al. (2017) used the iris ratio as a fixed measurement to quantify the growth of other landmarks. Although the two studies are different (Table 4), both are useful in predicting growth in juvenile faces.

Table 4: Machado et al. (2017) Vs Farkas (1994)

	Machado et al. (2017)	Farkas and Hreczko (1994)
Subject's age	6-22 years	0-25 years
Data	Longitudinal and cross-sectional	Cross-sectional
Population	Brazilian	North American
Sex	Non-specific sex	Male and Female

Machado et al. (2017) demonstrated the cumulative growth from 6-22 years of age as a percentage using the iris as a fixed reference. However, Farkas and Hreczko (1994) did not measure the iris diameter or the growth of the pupillary distance, therefore 8 landmarks were compared as shown in Table 5, and the difference between the two studies is shown in Table 6.

Table 5: Eight facial measurements from Farkas and Hreczko (1994)

	Measurement at age 6 years (mm)			Measurement at age 19-25 years (mm)			Growth Percentage (%)
	Male	Female	Mean	Male	Female	Mean	Male & Female
n-sn	40.1	39.3	39.7	54.8	50.6	52.7	32.75
ch-ch	41.7	41.2	41.45	54.5	50.2	52.35	26.30
n-gn	98.5	95.7	97.1	124.7	111.4	118.05	21.58
zy-zy	114.9	113.4	114.15	139.1	130.0	134.55	17.87
al-al	28.6	27.8	28.2	34.9	31.4	33.15	17.55
sn-gn	61.4	58.8	60.1	72.7	64.3	68.5	13.98
ex-ex	80.0	77.8	78.9	91.2	87.8	89.5	13.43
en-en	30.6	29.8	30.2	33.3	31.8	32.55	7.78

Table 6: Comparing facial anthropometric measurements between Machado et al. (2017) and Farkas & Hreczko (1994)

Order	1	2	3	4	5	6	7	8
Machado et al. (2017)	sn-gn	ch-ch	n-gn	n-sn	al-al	ec-ec	zy-zy	en-en
	28.80%	26.31%	26.13%	22.96%	21.15%	14.22%	13.63%	12.07%
Farkas and Hreczko (1994)	n-sn	ch-ch	n-gn	zy-zy	al-al	sn-gn	ex-ex	en-en
	32.75%	26.30%	21.58%	17.87%	17.55%	13.98%	13.43%	7.78%

**Exocanthion (ex)/Ectocanthion (ec) are the same landmark

It is interesting to see the order difference shown in Table 6, most noticeable with the height of the nose (n-sn), the height of the lower face (sn-gn) and the width of the face (zy-zy). It is uncertain what caused the difference between the two studies and perhaps it is the focal length of the camera and the 2D to 3D translation. The difference may also be due to population and face type, or the possibility of error in landmark placement especially at the zygion (zy). It is most likely to be a combination of all these factors.

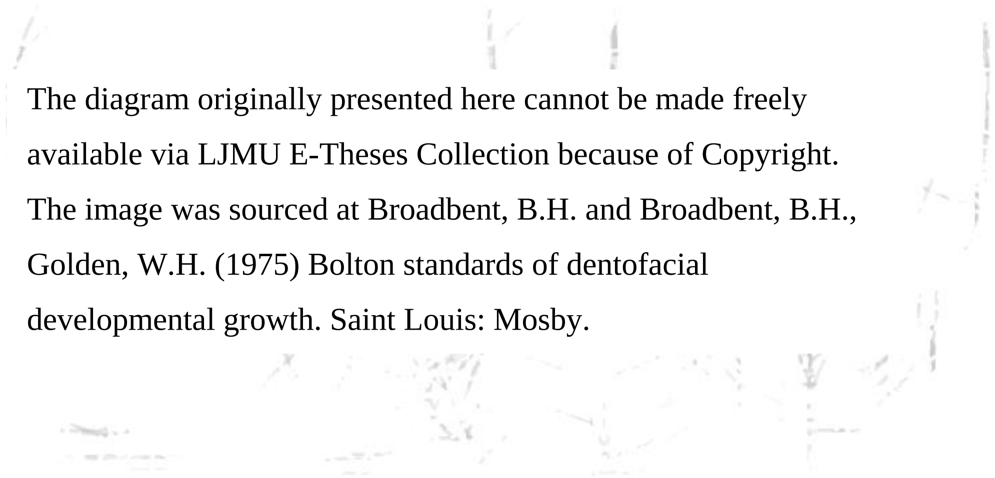
1.3.2.4 Bolton Standards (1975)

Bolton standards are averaged facial templates of Caucasian children derived from 22,000 cephalometric radiographs of 5,000 individuals between age 0-21 years old (Broadbent et al., 1975). This was a longitudinal study that began in the 1930s documenting the facial growth of healthy individuals with normal developing occlusion (Figure 18), and it was developed for orthodontic researchers to compare optimum facial and dental development growth (Broadbent et al., 1975). Averaged facial template transparencies were produced in frontal view for every age from 3-18 years old, and in lateral from 1-18 years old. These frontal templates could be used as a guide in age progression to estimate facial growth in frontal images of children, but these nonspecific growth templates are not ideal compared to gender-specific models (Erickson et al., 2016).

Young faces are similar between sexes until puberty, at around age 12 years, when the female face reaches maturity and the male face continues to grow into the early 20s (Kuroda et al., 2013). Each individual will have slight differences in the duration and timing of the pubertal growth spurt and the Bolton standards average these differences, resulting in a smooth incremental pattern in facial growth (Broadbent et al., 1975). The duration and timing of growth spurts in relation to changes in the facial pattern is also something an age progression cannot predict, and the Bolton standard is unlikely to be an accurate representation of puberty-related facial changes.

<u>The potential limitations to the implementation of the Bolton standards include:</u>

1. Population-specific (North American Caucasian)
2. Templates not sex-specific
3. These templates represent healthy children with normal developing occlusion, and therefore may not be applicable to children of other dentition classifications.
4. Growth rate varies between different individuals and sexes (Chronological versus Biological age), and individual differences in growth spurts will vary from the Bolton standard.
5. Depictions work best on frontal images, which are not always available - camera distortion of the original photographs could affect the alignment of templates.



The diagram originally presented here cannot be made freely available via LJMU E-Theses Collection because of Copyright. The image was sourced at Broadbent, B.H. and Broadbent, B.H., Golden, W.H. (1975) Bolton standards of dentofacial developmental growth. Saint Louis: Mosby.

<u>Figure 18: Bolton standards for juvenile craniofacial growth (Broadbent et al., 1975)</u>

1.3.3 <u>Morphing age progressions</u>

Lampinen et al. (2015) addressed the level of inter-artist subjectivity in age progression methods by morphing together images created by different artists. By averaging the progressions together, the agreement on certain changes could be emphasised thus minimising the differences. The authors showed that most age progressions with a large age range (age 5 – 20) were rated as dissimilar to the target photo, and age progressions from a younger age (age 5 - 12) showed a greater inter-artist variation when compared to an older

(12 - 20 years old) progression. These findings are supported by Erickson et al. (2016). It is logical to assume that there will be more variation in growth-related changes to the face with a wider age gap, and growth-related changes to the face are less variable as a child gets older.

1.3.4 Limitations with age progression

This section discusses the different aspects of limitation in age progression including:

- ‘Physical’ limitation
- ‘Psychological’ limitation
- ‘Recovery’ limitation

1.3.4.1 Physical limitation: Facial growth

For an age progression to be effective, the images have to represent what the child currently looks like and be able to provide improvements over an outdated photograph (Lampinen et al., 2010). Understanding the changes in facial growth from childhood to adulthood is key to developing or using any age progression methods or tools. The use of averaged growth pattern in an age progression may be an inaccurate representation of the child, as developmental rates vary between individuals even within a population, and these variations will introduce errors into age progression techniques (Lampinen et al., 2010).

Research related to the facial growth of children is well studied and appears in orthodontic related literature. A larger amount of longitudinal data following the growth of children has been established to gain an understanding of the factors affecting growth (Bishara et al., 1984; Bjork, 1963; Hans et al., 1994; Kau and Richmond, 2008; Ochoa and Nanda, 2004). Growth studies using longitudinal data are able to document information related to individual variation and this will have an advantage over cross-sectional data (Moss, 1964). Standards and methods of treatments were then developed for the orthodontist to achieve optimal results for different patients, and these standards were used to describe mean trends and not for predicting individual changes, as these changes vary within the same growth period and between sexes (Bishara et al., 1984).

Different body systems can have different maturation rates, and the difference in facial growth pattern is interlinked between the developments of organs within the head, the airway, the oral region and the basicranium[1] (Enlow and Hans, 1996; Gill and Naini, 2012). The head shape is determined by the neurocranium[2] configuration, which in turn will have an influence on face types (Enlow and Hans, 1996; Gill and Naini, 2012). Face shape will also be influenced by developmental factors related to the airway, mastication, dentition and occlusion (Franco et al., 2013). Although the distances between the eyes and the width of the nasal bridge remain relatively similar during growth, the eyes will appear to be closer together in relation to the vertical facial dimension at the cheekbones, and the nose increasing in size and height (Enlow and Hans, 1996). With changes to mastication and dentition during growth, the gonial region of the mandible extends laterally from the medial side of the cheekbone. This changes the v-shaped child mandible to a more 'squared' adult appearance (Enlow and Hans, 1996). The greatest influential factor for the face is perhaps the nasal area, in comparison to adults, young children and infants tend to have a lower nasal bridge with a 'pug-like' (upturned) nose (Enlow and Hans, 1996). During growth, the mid-face expands as a result of the changes in the anatomical positioning of the airway relative to lung capacity and body size, the male face tends to have a wider and longer nasal region to accommodate for a larger airway capacity (Gill and Naini, 2012; Kuroda et al., 2013).

Regardless of the difference in the head-form, face-form or sex, the prepubertal head and face are more brachycephalic, when the nasal region, dentition, jaw (mastication), and airway are not as developed as the neural component (i.e. brain) (Enlow and Hans, 1996). Facial growth is an equilibrium between functional and structural components, when the difference in growth across the ages are compared (e.g. the Bolton standard), this produces a model illustrating a forward and downward expansion seen in many studies (Enlow and Hans, 1996). Enlow and Hans (1996) & Kuroda et al. (2013) described two main head/face types along with three types of facial profiles (Table 7 & Figure 19). Ranges do occur, intermediate head-shapes are described as mesocephalic, and mesoprosopic for intermediate face-shapes. Head-shape does not always correspond with the associated face-shape and Enlow and Hans (1996) described the head form Dinaric, characterised by a brachycephalic head shape with a leptoprosopic face shape. Facial variations within and between populations are vast and diverse, regional imbalances during the developmental process in facial growth is an unavoidable event (Enlow and Hans, 1996). This will lead to a difference

[1] *Basicranium: The inferior region of the skull*

[2] *Neurocranium: The part of the cranium enclosing the brain*

in face shapes and also asymmetries, therefore using average templates as a guidance in manual age progression (Figure 18) is problematic, especially when growth patterns vary between different individuals and populations.

Table 7: Head/face shapes (adapted from Enlow and Hans (1996) & Kuroda et al. (2013))

Head Shape		Facial Profiles (Fig. 9)		
Dolichocephalic (narrow and long)	Brachycephalic (wider and rounder)	Orthognathic (A)	Retrognathic (B&C)	Prognathic (D)
Corresponding Face Shape		Straight-jawed	Retruding chin (most common)	Bold lower jaw and chin
Leptoprosopic (long and thin)	Euryprosopic (round and broad)			

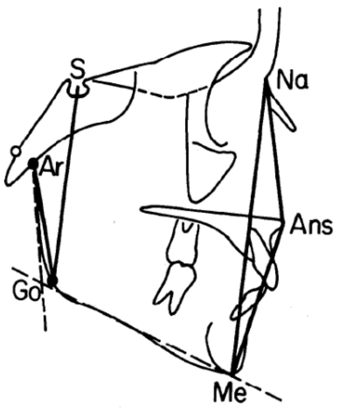
The diagram originally presented here cannot be made freely available via LJMU E-Theses Collection because of Copyright. The image was sourced at Kuroda, S., Schmittbuhl, S. and Nanci, A. (2013) Facial Growth and Development, in: Nanci, A. (Ed.), Ten Cate’s Oral Histology: Development, Structure, and Function. Missouri: Elsevier Inc., pp. 328–336.

Figure 19: Facial Profiles (Kuroda et al., 2013)

Children can exhibit different growth rates in the lower face and the upper face (Ligthelm-Bakker et al., 1992). Fields et al. (1984) suggested that the differences in the lower face height between long and short face children were related to the morphology of the mandible, where long-face children tended to have a larger gonial angle, greater dentoalveolar component, more intermaxillary space, and a greater posterior upper and lower dental height. Blanchette et al. (1996) observed that growth in the lower anterior vertical facial height in long faces were nearly twice as great, compared to the shorter faces.

Different face types can exhibit different growth patterns (Sassouni and Nanda, 1964). In adults, long-faces tend to be more retrognathic (Enlow and Hans, 1996) with a greater anterior lower face height (Fields et al., 1984), whereas a brachycephalic face tends to have a straighter or concaved profile (Enlow and Hans, 1996). These difference in facial morphology could be related to the shape of the dental arch (Rakosi et al., 1993), the difference in growth between the different head and face shape is particularly important for orthodontics to plan treatments (Enlow and Hans, 2008). With the palatal size difference between the different head and face shapes, broad faces usually receive expansion treatment, and extraction treatment for long faces (Rakosi et al., 1993). Dajani (2008) produced regression equations models to predict nasomaxillary growth, and these are beneficial in planning treatment. Research in this field is valuable for age progression research in predicting facial growth.

The literature describes face shapes as long or short using linear measurements of the total, upper and lower face height marked with various cephalopmetric landmarks such as Nasion (N), anterior nasal spine (ANS) and menton (Me) (Figure 21) (Blanchette et al., 1996; Fields et al., 1984). These are usually measured on lateral cephalograms. However, the literature uses the Face Index, also known as the Prosopic Index to describe and measure face shapes (Leptoprosopic / Euryprosopic) (Figure 20). Face shape is measured [(Face length/Face width) * 100], and this has been widely used to study facial variations between and within different populations (Hossain et al., 2011; Raji et al., 2010; Shah et al., 2015; Torres-Restrepo et al., 2014). Literature has defined Face height (N-Gn) and bizygomatic width (Zy-Zy) (Farkas et al., 1994; Franco et al., 2013) and most of the literature also measures the head shape (Dolichocephalic / Brachycephalic) using the Cranial Index [(Head width/Head Length) * 100], defining Head Width (Eu-Eu) and Head Length (G-Op) (Torres-Restrepo et al., 2014).



Morphologic face height (TAFH) = N-Me
Upper anterior face height (UAFH) = N-ANS

Figure 21: Face Height Measurements (Nanda, 1988)

© 1988, with permission from Elsevier

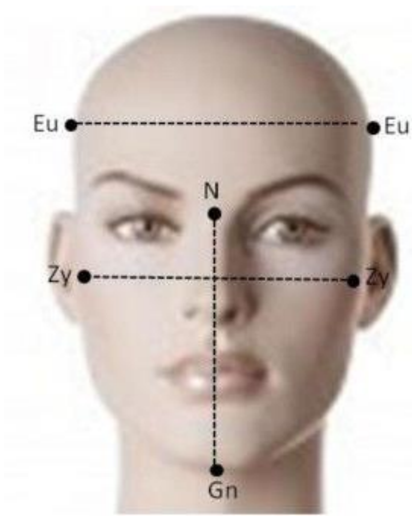


Figure 20: Facial Index

(© Torres-Restrepo et al., 2014)

Euryprosopic (Brachycephalic) face types will appear more juvenile resembling the wide and short facial configuration of a child, where a dolichocephalic face will appear to be more mature (Enlow and Hans, 1996). Could this suggest age progression of a child with a euryposopic face form would be easier to predict? Is facial growth population specific or face type specific?

1.3.4.2 <u>Physical limitation: Images</u>

There are three physical limitations relating to image use. Firstly, the quality of the reference/source photo; secondly, the quality of the age progression; and thirdly, the quality of the target images for matching using computer algorithms.

Challenges in pose, illumination and expression of the image have been an area of interest in machine-based facial recognition, and this is even more challenging with the change in shape and texture of the face related to ageing in children (Chellappa et al., 2010). The source image will influence the quality of the age progression (Lanitis and Tsapatsoulis, 2016) and in most cases, the reference (original) images will not be of ‘studio-posed’ quality, with unconstrained facial expression and head position making measurements and proportional predictions difficult (Farkas et al., 1994; Lanitis and Tsapatsoulis, 2016). If available, frontal images should be used as the basis for an age progression (Lanitis and Tsapatsoulis, 2016).

Manual age progression is a subjective method involving a high level of artistic judgment, and the quality of the likeness produced can vary between different artists (Frowd et al., 2014). There are currently no standardised methods or training (Figure 14), and depiction of the same individual can vary in style depending on the source material and personal judgment (Erickson et al., 2016). Erickson et al. (2016) suggested that there is a correlation between the experience of the artist and the similarities between the age progression and the target image.

The quality of the target image can also limit the match rate using FRS. In terms of indecent images of children, images are likely to vary in illumination, pose and facial expression. These limitations are variable factors that are difficult to control and problematic for identification.

The minimum interocular distance on a passport specification is 120 pixels (ISO/IEC 19794-5). Resolution of an image can affect the performance of an FRS (Hennings-Yeomans et al., 2008) and Boom et al. (2006) suggested a resolution of 32 X 32 pixels as optimal. Undoubtedly, image resolution will have an effect, but the 'optimal' may vary between different FRS and environmental conditions. For example, Grother et al. (2017) suggested that the optimal resolution for identification from turnstile video clips is between 20-55 pixels. Recognition using standardised passport images is certainly different to recognition in the wild, and factors such as noise, blurring, pixels and brightness will affect the algorithm (Dodge and Karam, 2016; Grm et al., 2017).

Research has incorporated different methods to improve super-resolution for face recognition (Hennings-Yeomans et al., 2008a, 2008b; Kong et al., 2013; Lin et al., 2007; Wheeler et al., 2007). Super-resolution/reconstruction is used to enhance low-resolution poor surveillance footage, although it is able to generate a higher resolution from a low-resolution image, this process often generates distortion and artefacts (Hennings-Yeomans et al., 2008b; Lin et al., 2007).

Photography is subjected to distortion including optical distortion from the camera lens, and perspective distortion from subject-to-camera distance (Mansurov, 2017a). Perspective distortion can impair facial recognition in human perception (Liu and Chaudhuri, 2003; Liu and Ward, 2006), thus many researchers have developed methods to estimate subject to camera distance or even to correct such effect (Gallagher, 2002; Lades et al., 1993; Mansurov, 2017b; Stephan, 2015; Wu et al., 2013). It is known that camera distortion can affect machine-based facial recognition (Lin, 2000), but how does perspective distortion affect facial recognition in the wild? Certainly, this will have an effect on the methods applied to age progressions, especially when performing facial anthropometric measurements on a photograph.

1.3.4.3 <u>Psychological limitation</u>

The primary purpose of an age progression is to recover missing children by triggering recognition. However, the reliability of using age progression depictions has been questioned by psychologists (Charman and Carol, 2012; Lampinen et al., 2012a, 2015).

Retrospective person memory (encountered the child before publicity) and prospective person memory (encountering the child after publicity) plays a significant role in different types of identification (Lampinen et al., 2010). Lampinen et al. (2012a, b) examined prospective person memory in the identification of a missing child, and this refers to actively looking for a missing individual. Retrospective person memory was also addressed; this refers to a situation where the target is identified (through posters or adverts) by remembering a face from the past. Lampinen et al. (2012a) presented three different types of ‘missing children’ images for participants to identify. Participants were randomly assigned to view either an outdated image (age 7), a current image (age 12 different to the image pool), or an age-progressed image (simulated age 12). Similar procedures were repeated in Lampinen et al. (2012b), where the age-progressed image was viewed alongside an outdated image. A forensic case was also used to test for inter-artist variability. Lampinen et al. (2012a, b) concluded that age-progressed images did not appear to be more useful than outdated images, and Lampinen et al. (2012b) added that age-progressed images could be prone to conservative response bias, which may result in the reduction of investigative leads.

Charman and Carol (2012) compared the recognition rates between the original out-dated image, the age-progressed image generated by a machine-based system (APRIL), and both combined. The authors suggested that an age progression image could distract from the true target and therefore increase the likelihood of a mis-identification. Participants performed worse when compared to the group who viewed the outdated images only, and this suggested that there could be a psychological effect that harms target recognition.

Instead of testing recognition, Erickson et al. (2016) asked participants to give a similarity rating on a Likert scale between the age progression and a photograph at the target age. This type of methodology could perhaps evaluate the reliability of the method of age progression (Charman and Carol, 2012; Lampinen et al., 2012a, b) by measuring validity. Most studies used unfamiliar face recognition tests carried out by university student participants. With research suggesting a difference in recognition rate between familiar and unfamiliar face recognition (Burton et al., 2015; Ellis et al., 1979; Natu and O'Toole, 2011), it is difficult to gauge the application of such testing. In a forensic context, the missing child could be both familiar and unfamiliar face.

Familiar face recognition is more resistant to variables such as image distortion (Burton et al., 2015), viewpoint, context, lighting, and expression (Johnston and Edmonds, 2009). Where these changes could make unfamiliar face recognition difficult (Johnston and Edmonds, 2009), could this have an effect on the way age progressions are presented to the public? Psychology research suggests that distortion to a facial image provides no reduction in familiar face recognition (Burton et al., 2015) and this raises a question about the effect of changes in spatial configuration for an age progression. Are age progressions unrecognisable because of the 'growth' related manipulation to the image, or is it caused by inaccuracy in texture or external features such as hairstyle?

Perhaps recognition is challenging due to inaccuracy in both shape (configural) and textural (non-configural) changes. External features such as hair colour and style could be changed easily, the greater the timespan of an age progression, the more variable the external features will be. Therefore it is logical to think that the estimation of external features will become more inaccurate as the timespan increases.

Concealing the external features of an age progression by cropping the images showed no significant difference in recognition (Lampinen et al., 2015). Concealing what is unknown could be beneficial in forensic settings, especially when the environmental condition of the child is unknown. Face perception can be sensitive to changes such as hair colour in Caucasians (Abudarham and Yovel, 2016). Concealing hair does not generally affect unfamiliar face recognition, but a change in hairstyle can affect recognition accuracy (Erickson et al., 2016). The process of unfamiliar face recognition can be affected by simple changes in appearance, which may cause disruption in recognition (Toseeb et al., 2012). Because both unfamiliar and familiar recognition is possible using age progression images, with evidence suggesting that the blurring of external features is beneficial to facial construction in recognition memory (Frowd et al., 2012). It would be logical to suggest this could also be beneficial for the final presentation of age progressions.

However, images used for human recognition should be treated differently to images used in a machine-based system. An image enhanced to optimise a facial recognition system, may not necessarily look 'good' or perform well in human recognition (FISWG, 2016). With the vast amount of data received related to indecent images of children, human recognition may be an ineffective and psychologically tiring method to pursue. The current machine-based age progression methods described above are still in early stages, where they are not well tested with no reported accuracies. However, even with the success stories for human recognition in the recovery of missing children using age progression (Goldman, 2009), should age progression be avoided?

1.3.4.4 Recovery limitation

Most recovery methods of missing children focus on publicising the identity of the missing child in hope for the public to report and contact authorities, and the success rate is dependent on the factors illustrated in Figure 22 (Lampinen et al., 2010):

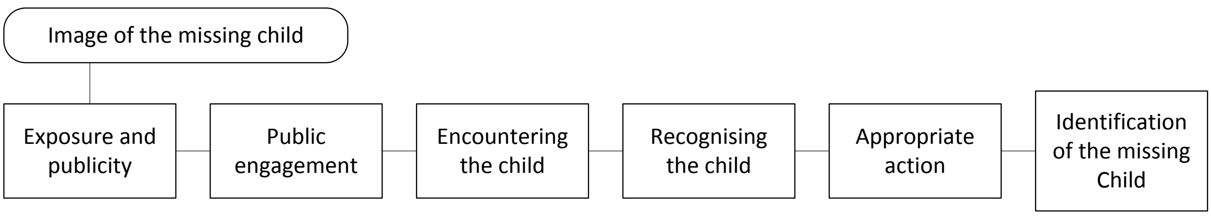


Figure 22: Recovery stages relating to a missing child (Adapted from Lampinen et al. (2010))

All of the stages shown in Figure 22 are the steps leading to the identification of the child; each step decreases the probability of identifying the child. Therefore, with a low success rate, disseminating the child's photograph is crucial to increase the odds of identifying the child. Lampinen et al. (2010) suggested that although people do care about missing children, images shown in public places such as a supermarket are often ignored. Therefore the authors suggested placing advertisements at or near the location of the sales point as an improved method to increase customer attention; this is also known as 'point of purchase advertising' where the identification rate performed significantly better than chance in comparison to placing the advertisement at the exit. Lampinen et al. (2012c) also explored the relationship between the number of targets (posters) used related to prospective person memory. The authors reported a slight increase in false positive identification when more targets (posters) were introduced. Lampinen and colleagues aimed to improve campaign designs, and concluded that the use of a large number of posters in current practice is acceptable. However, overuse can decrease the effectiveness of missing person alerts (Lampinen and Moore, 2016). With the vast number of children going missing along with the increasing displacements of populations and an overload of media information, human recognition may not be an effective identification method. This research thus focuses on the ability of machine-based methods for the recognition of children's faces over time.

2 Methodology

Chapter 2 is divided into two sections (Experiment 1 and 2) and each experiment has a different methodology designed to document the different approaches followed by the results.

1. Experiment 1: Group tagging: Age group and age gap versus automated recognition
2. Experiment 2: Age-progression
    a. 2A: A guided method for digital manual age progression
    b. 2B: Application of conditions for machine-based face recognition
    c. 2C: Manual age progression versus machine-based study

Experiment 1 aims to examine if group tagging is more beneficial for facial recognition across different ages of the same individual. Images of the same individual with a variety of ages were tested using Google Picasa, a facial recognition freeware. The recognition rate between different ages and the age gap were compared. The results from Experiment 1 will establish the age range necessary for age progression work in Experiment 2.

Experiment 2 compares the similarities between age progression images and the target images using FRS and manual image comparison. 2A addresses a guided manual age progression method developed using various growth studies to address whether different conditions affect FRS. 2B applies blurring, resolution reduction, cropping and decolourisation to the depictions from 2A to evaluate recognition, 2C compares manual age progression with a machine-based process.

2.1 Key objectives

Experiment 1

1. Explore the benefits and disadvantages of multi-source database
2. Compare recognition rates for different ages
3. Identify the optimal age gap for facial recognition of children using age progression images

Experiment 2

4. Evaluate the inter-observer error of the manual age progression method
5. Establish how recognition is affected by age and sex with the increased age gap
6. Explore how different conditions can affect machine-based recognition of age progression images
7. Compare the recognition rate for out-dated images and age-progressed images with veridical images
8. Explore the effectiveness of the proposed manual age progression method
9. Compare the recognition rate between manual and machine-based age progression methods

Experiment 1 and 2

10. Explore the limitations of the methods
11. Based on objectives above, make suggestions for age progression work and methods in the identification of children's faces

2.2 Novelty of research

Improvements and acceptance of technology has increased the use of facial recognition software. Ferguson (2015) suggested that facial recognition systems (FRS) are better than humans at identifying children's faces and this research further explores if age group and age gap can affect FRS and establishes if age progression work could further improve the recognition rate.

At present, research exploring the reliability of age progression is limited. Although the general trend of recognition in relation to age gap had been previously discussed (Ling et al., 2010; Mullins, 2012; NCMEC, 2016), none have analysed these faces under a verification setting using a longitudinal dataset. Although facial growth studies of children are well established, few address the application for use in age progression (Bulut, 2010; Farkas et al., 1994; Ramanathan and Chellappa, 2006). This research utilises ideas from a few previous studies and establishes a new method for digital manual age progression.

No comprehensive work has previously explored how different age progression conditions can effect FRS, and no research has compared the performance of manual age progression with a machine-based system.

Manual age progression is currently utilised as an investigative tool in cases of long-term missing children (NCMEC, 2016). The outcome of this research will directly benefit those who practice in facial identification of children, especially in age progression work.

2.3 Challenges

- Inaccurate documentation on the age of the images within the database
- Low-quality photographs (resolution, lighting, distortion, occlusion etc.)
- Extreme poses
- The unknown in using a black box system (i.e. Google Picasa and Microsoft Face API)
- Growth studies used for manual age progressions are averages and population specific.

2.4 Dataset

To be able to conduct research related to age progression in children, it was essential to collect data documenting the growth of children's faces. The public database FG-NET was used, along with data gathered from other open-access media, such as YouTube.

The FG-NET ageing database contains 1002 unstandardised images from 82 different individuals between 0-69 years of age. The database was released in 2004 and it has since been used in research related to face recognition, age estimation and age progression (Panis and Lanitis, 2014). For this research, only faces below 18 years of age were considered. Results generated could be comparable to other researches. Not all subjects within FG-NET had sufficient images between the ages of 0-18, Figure 23 shows the frequency of images between 0-23 years old. 78 subjects (32F; 46M) had between 4 to 18 images.

27 FG-NET subjects (14F, 13M) were selected, to generate 80 progressions for use in Experiment 2B.

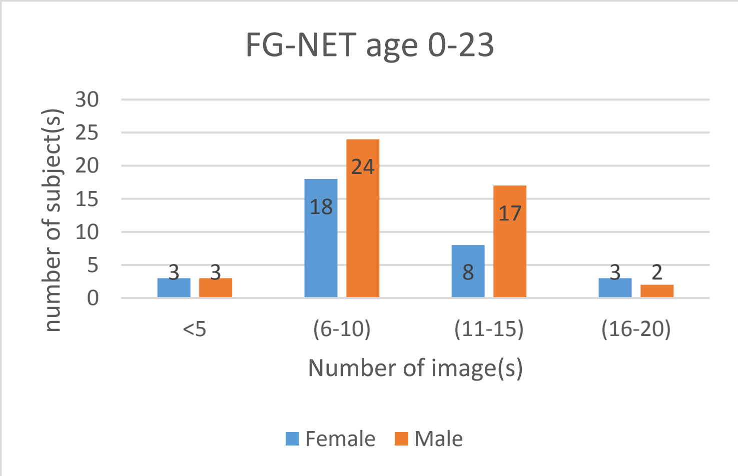


Figure 23: The age distribution of FG-NET images (Years)

With the limited existing public databases, this project also collected a child-ageing database with a view to providing this as an open-access research database.

2.4.1 Open-access Child Ageing Database

The adult public was asked to donate unstandardised images taken from the camera, phone, webcams etc. of their younger self, and these images were organised into different genders and ages.

The collection of this database was approved by the Liverpool John Moores University Research Ethics committee on the 30th Oct, 2015 [15/CMP/005]. All subjects provided informed consent for image use in the context of academic research and publication. All images will be copyrighted to Liverpool John Moores University; however, should these resources be used in published research all that is required is for the authors to credit the database.

2.4.2 Data collection option A: On-line upload system

- Participants could donate images online via the following link: [http://cmpproj.cms.livjm.ac.uk/faces/].
- Participants read the participant information and provided informed consent before image donation.
- There was a tick box to ensure participants had read through, agreed and understood the research.

- The uploaded images were stored on the LJMU server, where the images can only be accessed by the researcher and supervisor.
- Participation was anonymous, and the e-mail address provided for contact were stored as an identifying code.
- Participants specified the subject age in each photograph.

2.4.3 Data organisation

Each participant was asked to donate images of themselves with an optimal number of 10 images, but participants were welcome to donate more or fewer. They were also asked to provide a few photographs across other age groups. Their email address was automatically anonymised as participants upload their image onto the server via the webpage.

The images appeared in the format as follows: [randomised code]-**Af01Y**-

- [randomised code] remained the same when the same email address was entered
- 6 ethnic groups: W-white, B-Black, A-Asian, M-Middle eastern, X-Mixed ethnicity, O-Others
- Gender: Female, Male or others [i.e. f/m/o]
- Age of the photograph: [e.g. 01 is age 1]
- If participants have agreed for their image to be used in published work [Y(yes)/N(no)]

Figure 24 shows the frequency of images for all subjects from the Open-Access Child Aging Database. 26 subjects (19F; 7M). The number of images per individuals varies. Some images were distorted, as it was not captured by a scanner. Unfortunately, the dataset was very small. One subject was used for the inter observer error test in experiment 2A.

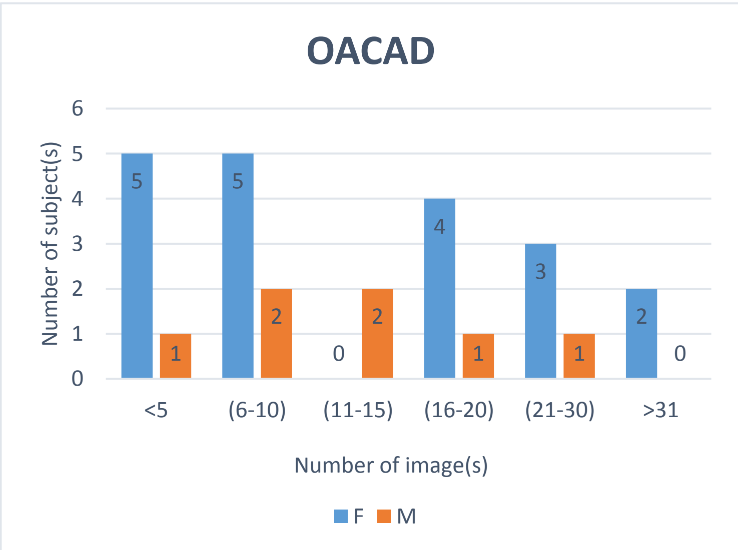


Figure 24: The age distribution of OACAD images (Years)

2.5 Method of recognition

This project used open-sourced facial recognition software to explore facial recognition accuracies. Picasa is a freeware originally created by Lifescape and acquired by Google Inc. in 2004 (Protalinski, 2016). The support for this freeware was discontinued on 15th March 2016 (Sabharwal, 2016).

This software allowed users to organise, share, view and edit digital photographs with functions such as tracking, tagging and facial recognition. With Miller et al. (2015) and Schroff et al. (2015) suggested that the facial recognition technology developed by Google was advantageous in comparison to other software, Picasa was chosen to be used as a tool to represent the facial recognition rate performed by a machine-based system. It is unknown how the algorithm was trained or set up, but other researchers had used Picasa as a tool to survey or compare between similar algorithms on performance, such as face detection and identification (Becker and Ortiz, 2013; Mazura et al., 2012; Qin et al., 2011; Zhu and Ramanan, 2012). It is important to note that the focus of this study is on the concept and limitation in using an algorithm to explore the relationship between juvenile ageing and face recognition.

Microsoft Cognitive Services (also known as Microsoft Project Oxford) contains a collection of cloud-based machine learning APIs (Application Programming Interface), and the Face API within was able to perform face detection, face verification, similar face search, face grouping and face identification with predictions to gender and age of the image. This API ran in Microsoft Visual Studio (C# application) and required an Azure account. The facial verification feature was able to return a confidence score between 0 and 1 on the identity match between two images. This feature was beneficial in comparing and analysing the recognition rate between out-dated images and the age progression in Experiment 2. Previous studies have implemented the Microsoft Face API (Bhuvaneshwari et al., 2017; Dehghan et al., 2017; Maheshwari and Nalini, 2017).

Both of these commercially available software represent machine-based methods, and any conclusions can only inform a recognition rate for a black box system. Research in FRS for the use in standardised photographs has almost reached its saturation, and there are many freely available FRS. Using a black box system in facial recognition studies had been used in previous research (Grother et al., 2010; Leonard, 2016).

2.5.1 Application of Google Picasa

Google Picasa was used when identifying faces from a large pool of faces for face verification using multiple source images as a group. Becker and Ortiz (2013) noted that Picasa could generate different results when the images were imported at the same time, compared to individually. By importing the images together, Picasa was able to consider the data more globally on the distribution of identities. This was one of the disadvantages of using a black-box system, as results can be inconsistent if the images were imported individually. Experiments using such systems require careful design.

- Images were collected into folders labelled with a specific individual code, with sub-folders separated into each age group
- The facial recognition function for the folders specific to the individual was enabled

In Picasa [Tools > Folder Manager > scan Always]

- One image at a near frontal view with minimal facial expression was selected at random to represent the individual

Select the image in Picasa, under [Add a name], input the candidate code

- Once the image was ‘tagged’, Picasa scans the photo library automatically
- Within the folder(s), certain photos were matched by Picasa as the same individual
- Each photograph within the folder was examined, and each match by Picasa was documented as a ‘Hit’ (positive identification)
- The age gap where an individual was no longer recognised by the software was evaluated

2.5.2 Application of the Microsoft Face API

The Microsoft Face API was able to generate a confidence score between two images, it was used for single image comparison in Experiment 2 of the study comparing different age-progressed images

- To run the program, an Azure account was acquired for a subscription key
- This program was run in Microsoft Visual Studio

3 Experiment 1: The effect of the multi-source image database

Experiment 1 used video clips from YouTube; the clips show a collection of videos or photos documenting the individual 'growing up' from a young age. These videos document the growth of a child's face and were unconstrained with a variety of facial expressions and head poses between the ages of 0-16 years old (FC001) (Hofmeester, 2015), 0-13 years old (FC002) (Ani Acopian, 2014) and ages 1-11 years old (MC001) (Hofmeester, 2014). The video quality between FC001 and MC001 were similar as they were taken by the same individual, in the same style; with two sibling subjects. The video quality of FC002 was in a different style composed of still images translated into a video; the resolution was also lower in comparison to the FC001 and MC001. No copyright infringement was intended under fair dealing for research, governed by Section 29 and 30 of the Copyright, designs and patents Act 1988[3].

3.1 Age group versus age gap using a multi-source image database

Subjects with longitudinal datasets (faces across a wide age range) were used to examine how the FRS recognises faces of the same individual across different ages. Video clips from YouTube documenting the facial growth of the same individual were analysed as 'jpg' files, and Images were separated into age groups (Table 8) for comparison and recognition rate between the different age groups were compared.

Table 8: Age groups utilised for longitudinal datasets

Group 1: Age <1	Group 6: Age 10-11
Group 2: Age 2-3	Group 7: Age 12-13
Group 3: Age 4-5	Group 8: Age 14-15
Group 4: Age 6-7	Group 9: Age 16-17
Group 5: Age 8-9	Group 10: Age 18-19

Each age group in each video clip contained between 163-651 images depending on the subject, and each subject was tested separately using Google Picasa. Image(s) within a certain age group of each individual were used as 'source-image(s)', which was also referred as a reference image or the original image.

[3] Copyright, Designs and Patents Act 1988 s.29 – 30.

- The ‘sources-image(s)’ were tagged as the subject
- Image(s) within other age groups were tested against the source
- Positive identification(s) were recorded and collectively produced a percentage of recognition rate between age groups

3.1.1 T1 Methodology (Single-source)

1. ‘Tagged 1’ was abbreviated to ‘T1’ meaning only one image was tagged as the source image, which was also referred to as “single-source”
2. One frontal or near-frontal facial image with minimal facial expression was selected at random (Tagged as the subject)
3. The number of images suggested as the subject within each age group was recorded
4. Each age was repeated twice with different source images
5. An average between the repeats was displayed as a percentage

3.1.2 T5 Methodology (Group tagging)

1. ‘Tagged 5’ was abbreviated to ‘T5’ meaning five images were tagged as source images, which is also referred to as ‘multi-source’ or ‘group tagging’
2. Five frontal or near-frontal facial images with minimal facial expression were selected at random (Tagged as the subject)
3. The number of images suggested as the subject within each age group was recorded
4. The results were displayed in percentages as 3D charts

3.1.3 Subjects

Each subject was evaluated using Google Picasa. Images were collected into folders separated into each age group. The facial recognition function for the folders specific to the individual was enabled (i.e. Tools > Folder Manager > scan Always). Analysing only one subject at a time, images from each age at a near frontal view with minimal facial expression were selected at random to represent the individual. A candidate code (i.e. FC001/FC002/MC001) was added to the image. Once the image was ‘tagged’, Picasa scanned the photo library automatically. Within the folder(s), certain photos were suggested by Picasa as the same individual. Each photograph within the folder was examined, and each match by Picasa was documented as a positive identification. The positive identifications were manually collated to give a percentage of recognition in each age group. This provided an indication of the age gap over which an individual is no longer recognised by the software.

The test was separated into Single-source (T1) and Multi-source (T5). In order to explore whether using more images at source is better, an average of two images from each age was selected to represent single-source (T1), and the two images were tested separately. For multi-source (T5), five images at each age were tagged for the multi-source (T5) tests. The setup of this test could also indicate the percentage of age at which the faces were identified.

3.2 Experiment 1: Results

The number of facial images extracted from the 'growing up' YouTube video is listed in Table 9 to 11 below. These videos were analysed into 25 JPEG/sec; depending on the length of the video, each age group consisted between 163-651 faces. Of these, 75/3341 images (FC001), 10/1862 images (FC002) and 162/3230 images (MC001) were not detected as faces. Faces that were recognised within the age group were recorded as a percentage of the detected face photographs.

Table 9: FC001 facial image data

FC001 (VIDEO)	age <1	age 2-3	age 4-5	age 6-7	age 8-9	age 10-11	age 12-13	age 14-15	Total
Images (n)	426	395	401	406	402	455	404	527	3416
Face	405	358	394	406	397	455	402	524	3341

Table 10: FC002 facial image data

FC002 (VIDEO)	age <1	age 2-3	age 4-5	age 6-7	age 8-9	age 10-13	Total
Images (n)	456	470	275	262	163	236	1862
Faces detected	454	468	274	260	163	233	1852

Table 11: MC001 facial image data

MC001 (VIDEO)	age <1	age 2-3	age 4-5	age 6-7	age 8-9	age 10-11	Total
Images (n)	586	651	526	523	608	336	3230
Faces	549	627	486	471	602	333	3068

Figure 25, Figure 27 and Figure 29 shows the single-source (T1) condition for the three subjects. Figure 26, Figure 28 and Figure 30 shows the multi-source (T5) condition for the three subjects. Each figure describes one condition for one individual. See Appendix 1 for the table representation.

FC001, FC002 and MC001 across all ages achieved a higher recognition rate when more images (T5) were used: [$t(58)$ -7.635, $p < 0.001$] (Table 12). Results from the single-source recognition suggest that two different images of the same individual at a near frontal view with similar 'minimal' expression generated different recognition rates, T1 was an averaged recognition score between two single images. The inconsistent recognition rate of T1 also indicates that a single image is not representative of the recognition rate of the subject.

Table 12: T-test between single images (T1) and multiple images (T5)

Independent Samples T-Test					
	Levene's Test for Equality of Variances		t-test for Equality of Means (Equal variance assumed)		
Subjects	F	Sig.	t	df	Sig. (2-tailed)
FC001 (Age 0-16 Years)	3.675	.065	-3.901	30	.001
FC002 (Age 0-10 Years)	2.062	.168	-5.149	18	.000
MC001 (Age 0-11 Years)	2.622	.121	-3.230	20	.004
Averaged (Age 0-10 Years)**	2.362	.130	-7.635	58	.000

*** Although FC001 have data outside age 10 years and above, only age 0-10 years old were taken into consideration to show the averaged recognition rate due to missing data in other subjects.*

3.2.1 Comparing the use of multiple images (T5) to single-source (T1):

When the selected sources-image(s) from each age was compared to the different age groups, FC001 had 128 age group comparisons up to 15 years old; FC002 had 60 age groups comparisons up to 9 years old; MC001 had 66 age group comparisons up to 10 years old. None of the single-source (T1) performed better than the multiple image (T5) condition; the results below documented the number of T5 age groups that performed 30% and 50% better when compared to T1.

- FC001: 32/128 (25%) age groups of the T5 (Figure 26) performed 30% better in comparison to T1 (Figure 25) of which 4/128 (3%) age groups of the T5 performed 50% better.
- FC002: 49/60 (82%) age groups of the T5 (Figure 28) performed 30% better in comparison to T1 (Figure 27); of which 38/60 (60%) age groups of the T5 performed 50% better.
- MC001: 12/66 (12%) age groups of the T5 (Figure 30) performed 30% better in comparison to T1 (Figure 29) and none performed higher than 50%.

The match rate for FC001 and MC001 were similar and also considerably lower when compared to FC002, which suggests that recognition rate will vary with different individuals or different photographic quality. Since T5 shows better performance, this demonstrates that group tagging multiple photographs of the same individual increases the recognition rate.

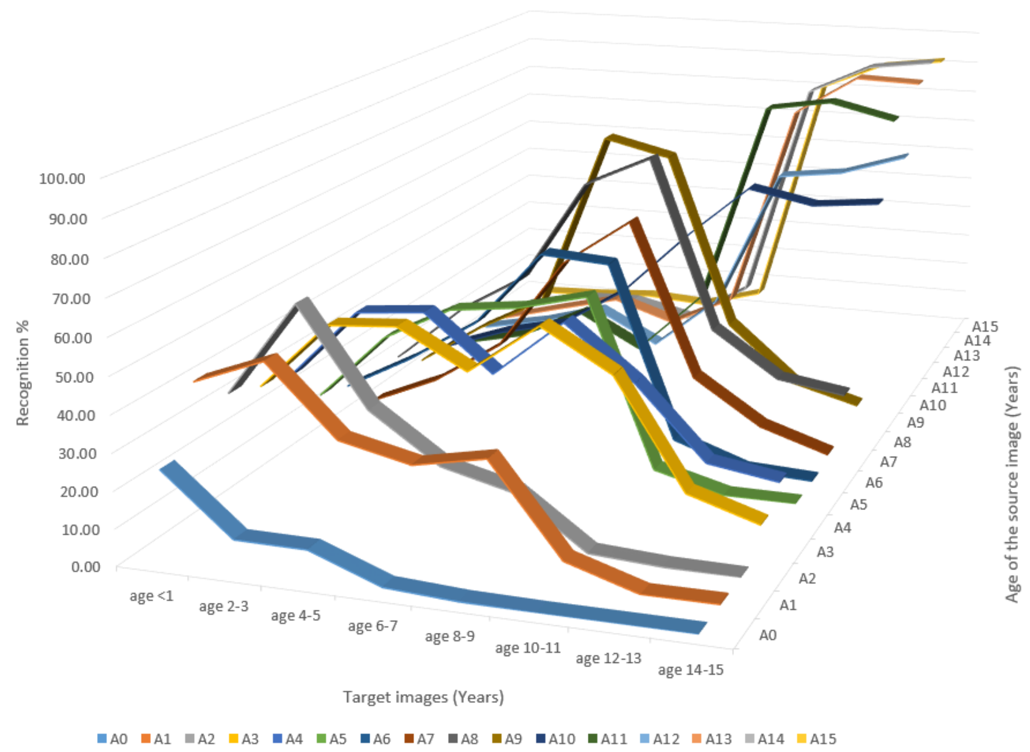


Figure 25: FC001 The recognition rate for each age using a single image (T1) compared to all images of the same individual (Target)

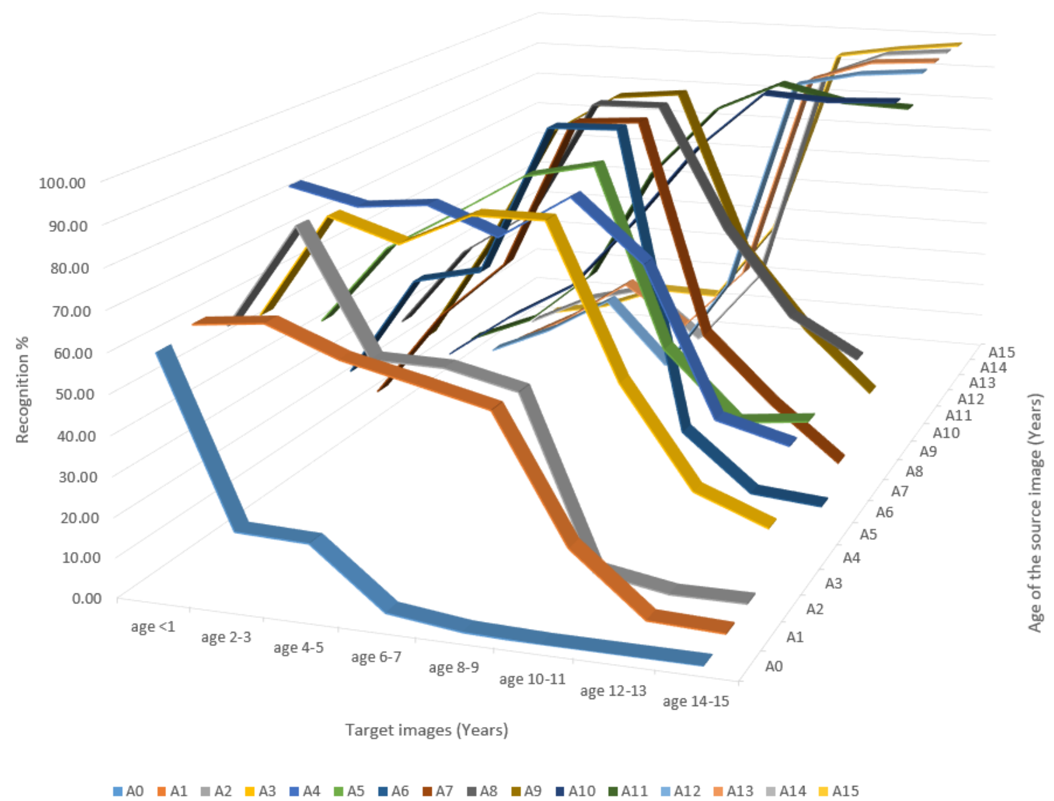


Figure 26: FC001 The recognition rate for each age using five images (T5) compared to all images of the same individual (Target)

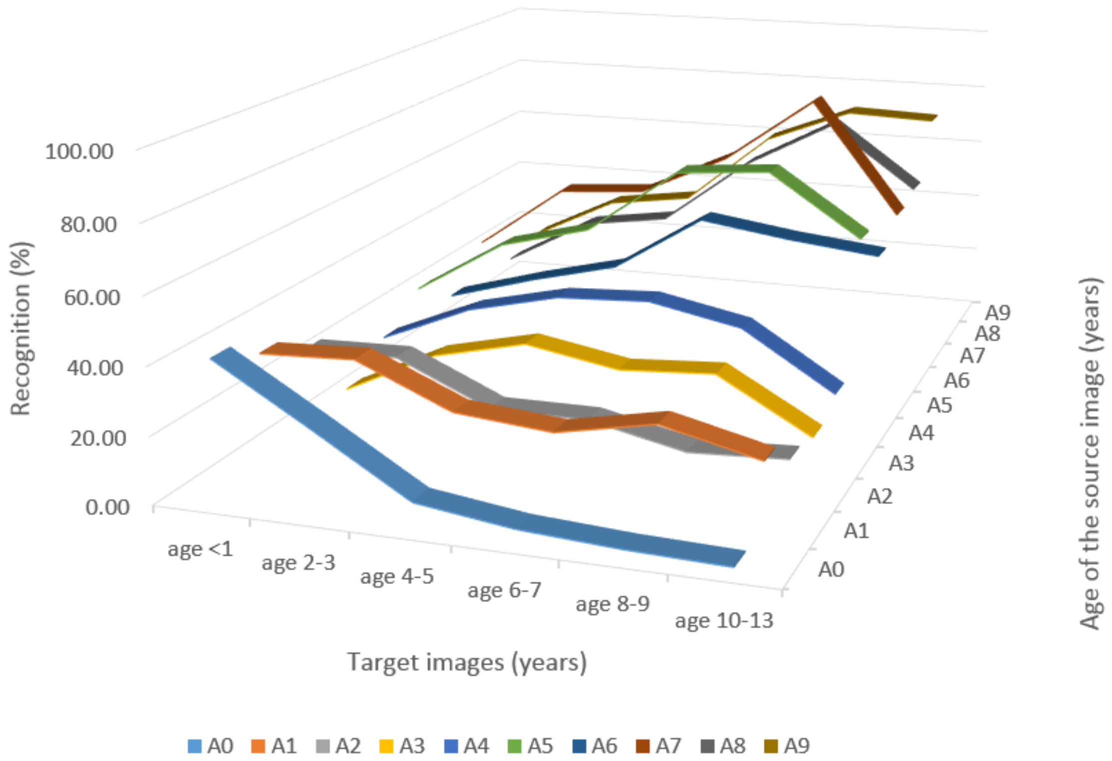


Figure 27: FC002 The recognition rate for each age using a single image (T1) compared to all images of the same individual (Target)

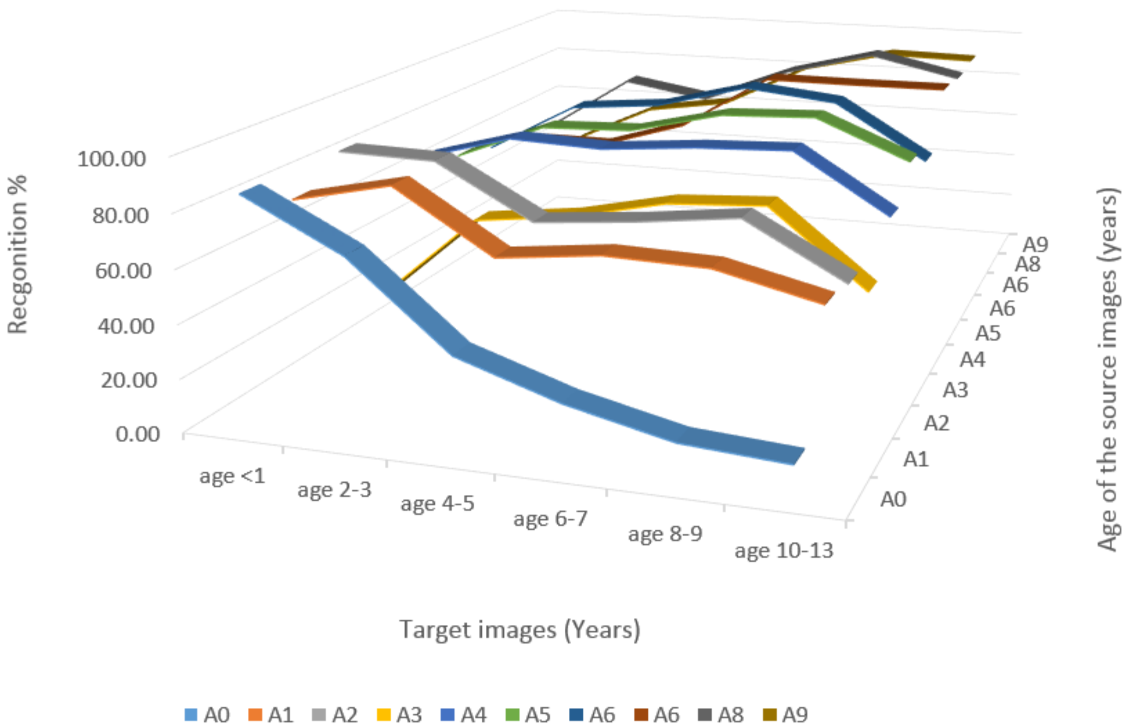


Figure 28: FC002 The recognition rate for each age using five images (T5) compared to all images of the same individual (Target)

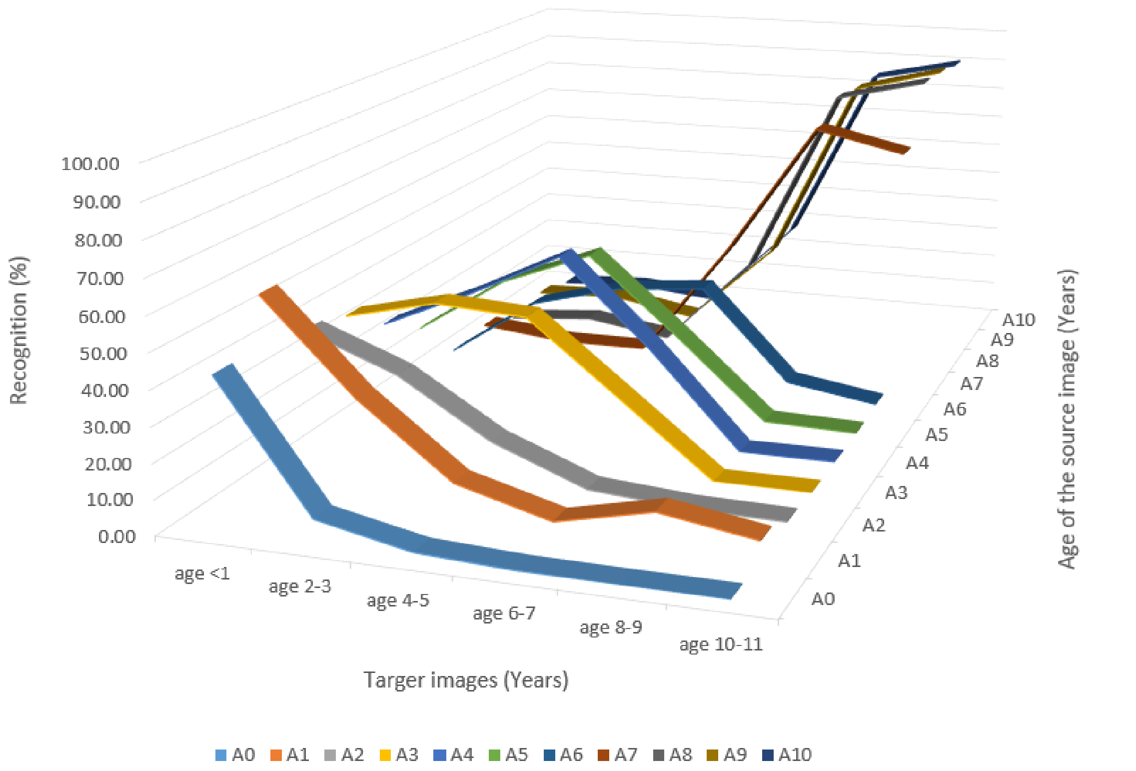


Figure 29: MC001 The recognition rate for each age using a single image (T1) compared to all images of the same individual (Target)

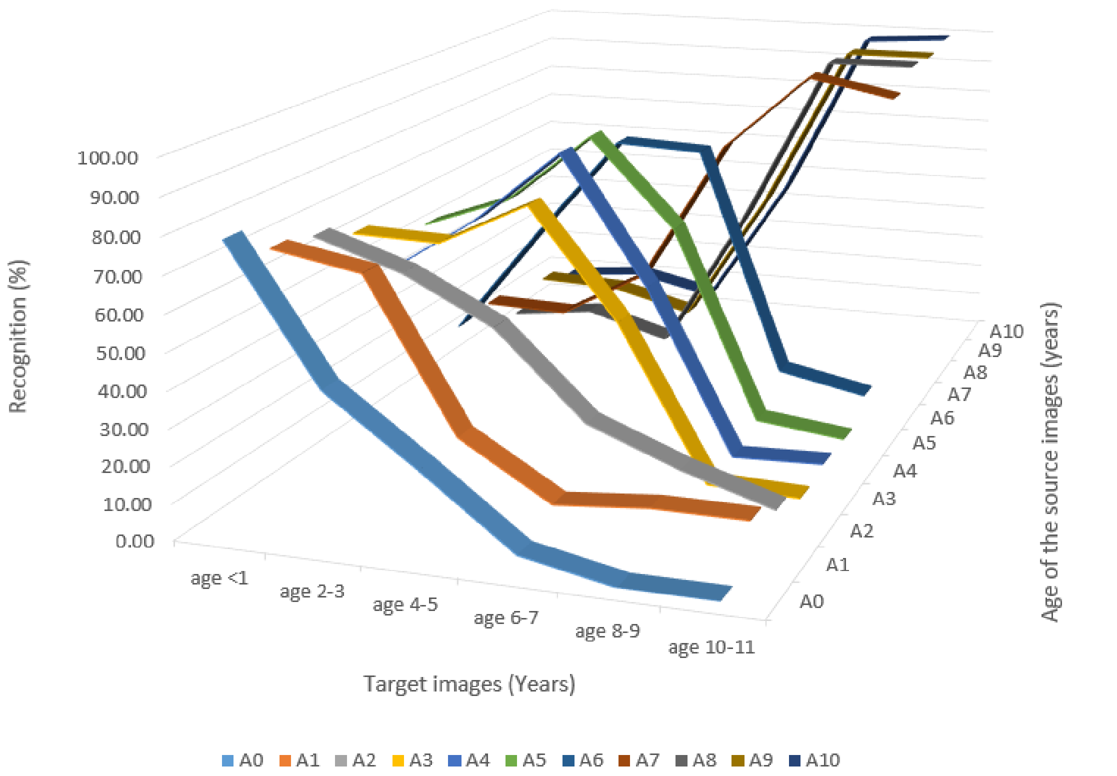


Figure 30: MC001 The recognition rate for each age using five images (T5) compared to all images of the same individual (Target)

3.2.2 T5 Recognition and original age

- FC001: from age 3 years, around 80% of the faces belonging to the same individual were recognised +2-6 years of the source age with exception to age 8 and 9.
- FC002: from age 4 years, around 80% of the faces belonging to the same individual were recognised +2-4 years of the source age.
- FC001 and FC002: At age 1 year and above, over a half of the faces were recognised across four other age groups.
- MC001: from age 7 years, around 80% of the faces belonging to the same individual were recognised +2-4 years of the source age.
- Recognition of MC001 was inconsistent across the age groups. This could suggest either the quality of the photographs varied below age 7 years, where the computer was unable to recognise most images as the same individual, or this could suggest the algorithm was unable to recognise the growth induced change to the face for this particular individual.
- MC001: At age 1 year and above, over a half of the faces were recognised within +2 years of the age at the source.

For FC001, FC002, and MC001, age 1 year and below (age <1) achieved poor recognition rate in other age groups (below 50%). This suggests that the extent of facial changes after age one is problematic for recognition using this algorithm. This part of the study provided a baseline on whether an age progression could be more useful.

It is also interesting to note the difference between FC002 when compared to FC001 and MC001. The distribution of recognition in relation to the age gap was more gradual with FC002.

The quality of the images within each age group could be represented by the percentage of recognition of the same age group. For example, T5 recognition from Age 2 against the age group ‘2-3 years’ for MC001 was around 60%; this was the lowest inter-recognition rate for MC001. If a benchmark of recognition rate below 80% was taken for the inter-recognition of the T5 condition, this includes age <1 year for FC001; age 4-5 for FC002; age <1 year, 2-3 years, 6-7 years for MC001. This inter-recognition rate could suggest that the validity of the target recognition rate for that particular age group will be less representative (i.e. a lower recognition rate across an age group could be caused by the inconsistency of facial images within that particular age group).

The recognition score from the nearby age groups could also indicate the image quality of that particular age group. If the neighbouring age group achieved a higher recognition score in comparison to the inter-recognition rate, this could indicate a more variable image set. Table 13 shows variable image set for T5 comparison:

Table 13: Inter-recognition rate with a lower recognition score in comparison to the neighbouring age groups

Subject	Variable image set
FC001	**A1**, A5, A6, A7, A13
FC002	A1, A3, **A4**, **A5**
MC001	**A2**, **A3**, **A6**, **A7**, A9

*Red indicates an inter-recognition rate below 80%

When the low inter-recognition rate was compared to the recognition rate of the neighbouring group, the group with 'variable' images are consistent. It is clear that the images of MC001 was more varied and less representative of the target recognition at A2, A3, A6 and A7. This suggests the image comparison for MC001 was more challenging.

To simplify a comparison between T1 and T5, the recognition rate of images in their own age group from all subjects were averaged (Figure 31). This assumes that recognition is most representative at the same age with the minimum changes in facial growth. The error bars are the range from all subjects, due to the different age range between the subjects, only FC001 had data for the age group 11-15 years; this age group was excluded from Figure 31.

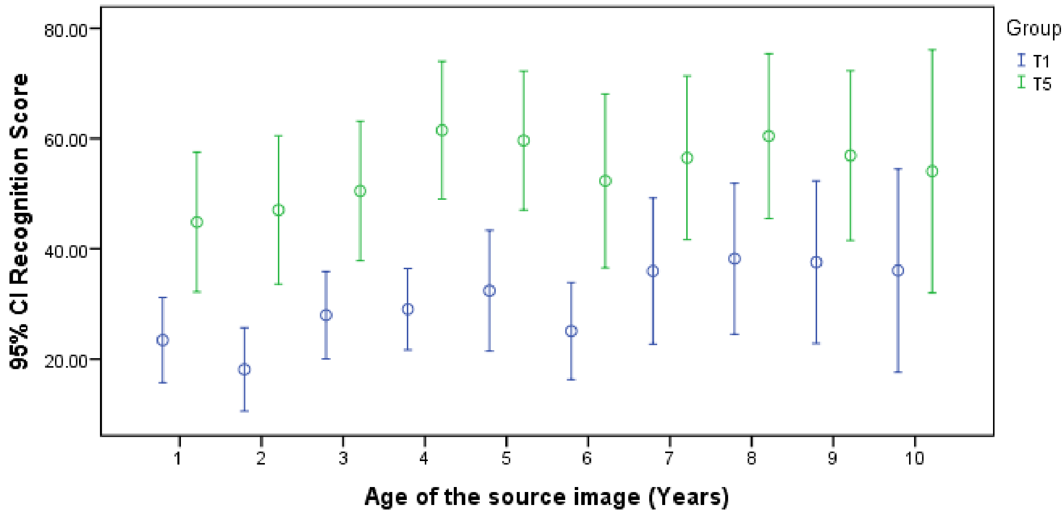


Figure 31: Averaged single image (T1) and multiple images (T5) recognition rate between all subjects in its own age group

4 Experiment 2: Age-progression

This chapter addressed the experiment focussed on the possibilities and limitations of age progression.

1. Experiment 2A: Testing a guided method for digital manual age progression
2. Experiment 2B: Evaluation of the conditions affecting machine-based face recognition
3. Experiment 2C: Comparison of manual age progression versus machine-based age progression

Using anthropometric growth studies, Experiment 2 introduced a guided method for digital manual age progression. This method described in Experiment 2A was used throughout this chapter. Based on previous research and results from Experiment 1, age progressions in Experiment 2B was carried out with a minimum of a 2-year gap between the subject age and the target age.

The ‘accuracy’ of an age progression was measured using machine-based face recognition. The confidence score generated by the Microsoft Face API was used throughout this chapter. With the availability of comparable data where studies have used the same FG-NET subjects, Experiment 2C compareed the performance of manual age progression to a machine-based method developed by Kemelmacher-Shlizerman et al. (2014). In Experiment 2C age progression images was matched to the image sets from the supplementary material of Kemelmacher-Shlizerman et al. (2014).

4.1 A guided method for digital manual age progression

4.1.1 Method of manual age progression

Using the FG-NET database limited the methodology as there were no accompanying photographs from family members. Having reference images of the biological parents and siblings around the target age of the missing child is optimal; without those images, more general reference images from other subjects of a similar age was recommended (Mullins, 2012; Taylor, 2000). The method of age progression followed the guidelines provided by Mullins (2012):

- Stretch the whole face by pulling down the area of the face just beneath the eyes: Elongate the lower 2/3 of the child's face to depict age-related changes
- The head remains roughly the same size after age 3 years
- Reshape the mouth and add darker shadows alongside the nose
- Addition of smile lines, remove baby fat and sharpening the angle of the lower jaw
- Not to alter the inner pattern of the ear, important for identification.
- A smile is unique, no drastic changes to the mouth and lips
- Thicken eyebrows and facial hair for male
- The facial characteristics of the missing child should remain recognisable, keeping features such as moles and scars
- If reference images were used, only take small portions from 5-6 different sources to ensures the individual of the reference image is not identifiable

Mullins (2012) also suggested altering the neck, clothing and hairstyle. However the present study explored FRS, so these features were not altered as they were not an area of interest. If dentition is visible, the guidelines suggested deciduous dentition from the original were depicted as permanent teeth (Mullins, 2012), but dentition is highly individualised and is often used in identification (Avon, 2004; Silva et al., 2008). Therefore, the lips were depicted closed so that teeth were not shown to avoid inaccuracies.

In addition to the method suggested by Mullins above, some growth trends from Farkas (1994) were taken into consideration in the method of age progression:

- The width of the head (eu-eu) matures around age 14 or 15 years
- The height of the head (v-n) matures around age 13 years
- Eye fissure height changes very little at around 1.3mm

The input image was first enhanced to improve the quality before conducting an age progression (Lanitis and Tsapatsoulis, 2016). The original images were altered as little as possible to retain facial characteristics (Farkas et al., 1994) and small portions from the reference material of other children were used (Mullins, 2012). The growth prediction was based on growth measurements from Farkas (1994) and the Bolton standards. Using the guidance by Farkas et al. (1994) and Machado et al. (2017), 11 facial measurements were selected as a guide to estimate facial growth (Table 14). The Bolton templates were also superimposed onto the images for guidance (see *Manual age progression*, p.36).

Farkas and Hreczko (1994) took physical measurements in millimetres and this can be difficult to translate into a photograph. As suggested by Machado et al. (2017) and Ronneburger et al. (2006), the iris diameter is a relatively stable measurement throughout growth, and this measurement can be set as a fixed reference. Unlike Machado et al. (2017) where the images are standardised, the quality of images can vary in resolution and subject-to-camera distance. By calculating the percentage of growth differences between two different ages (Table 16), this ratio can be translated into pixels measured in Photo editing software (Adobe Photoshop).

Table 14.1: Facial anthropometry (Amended from Farkas (1994))

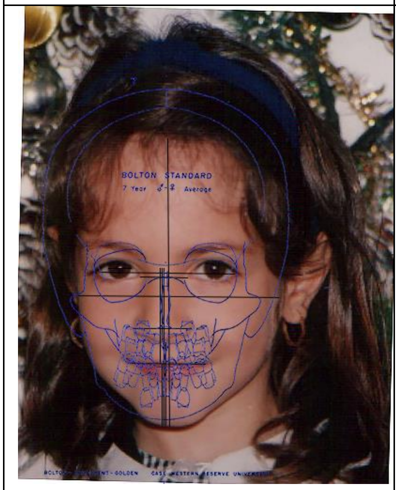
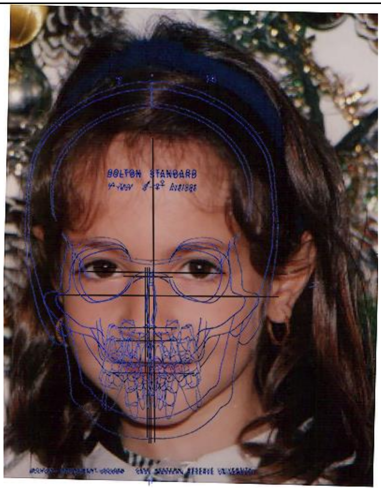
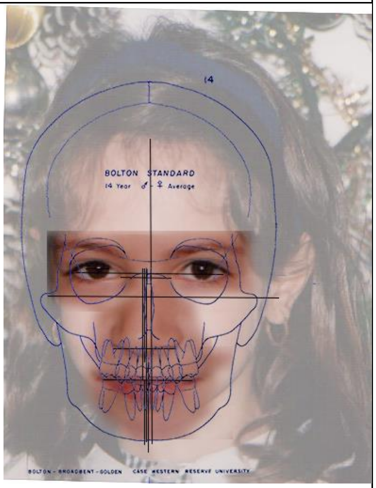
Male (mm)	zy-zy	ex-ex	en-en	al-al	ch-ch	n-sto	n-sn	n-gn	sn-gn	sto-gn	tr-gn
Age 1	96.7	76	27.3	26.5	34.8	49	30.9	80.6	49.9	31.9	143.6
Age 2	98.9	76.2	26.5	25.6	35.2	52.5	33.7	87.5	54.5	36.1	150.1
Age 3	101.4	77.5	27.2	26.1	36.7	54.3	35.3	88.5	55.2	35.6	153.4
Age 4	110.2	77.2	30.3	28.4	38.9	58.9	39.5	96.4	60.1	41.1	157.5
Age 5	111.8	78.7	30.8	28.9	40.7	58.6	38.9	96.7	60.3	42.2	155.4
Age 6	114.9	80	30.6	28.6	41.7	60	40.1	98.5	61.4	41.4	157.6
Age 7	116	79.2	30.2	28.8	42.7	60.4	41.4	99.5	61.1	42.4	161
Age 8	120.5	81.5	31.2	29.8	44.6	61.8	42.1	101.8	61.9	42.2	163.4
Age 9	121.8	82.9	31.7	29.4	45.5	62.3	43.7	102.7	61.7	42.4	163.8
Age 10	121.9	82.8	31.2	30.2	45.9	64.5	45	105.2	63.5	43.3	166.1
Age 11	125.7	85.2	32.6	30.1	46.4	65.4	45	107.1	56.3	44	169.5
Age 12	125.5	85.6	32	31.6	48.2	67.3	47.5	108.1	64.8	44.1	173.5
Age 13	128.5	86.8	32.8	32.4	49.1	68.3	48.8	111.6	66.5	45.7	175.4
Age 14	130.9	86.9	33.1	33.1	50.1	70	49.7	114.1	67.8	46.4	176.4
Age 15	133.5	89.4	33.7	34.2	51.8	73.3	51.9	119.1	70.6	47.8	184.8
Age 16	134.9	89.7	33.4	34	52.1	74.1	53	120.9	71.3	48.9	185
Age 17	139.1	90.7	33.9	34.8	53.5	74	53.2	120.9	70.8	48.5	184.1
Age 18	137.1	89.4	32.9	34.7	53.3	74	53	121.3	71.9	50.1	187.5
Age 19-25	139.1	91.2	33.3	34.9	54.5	76	54.8	124.7	72.6	50.7	187.2

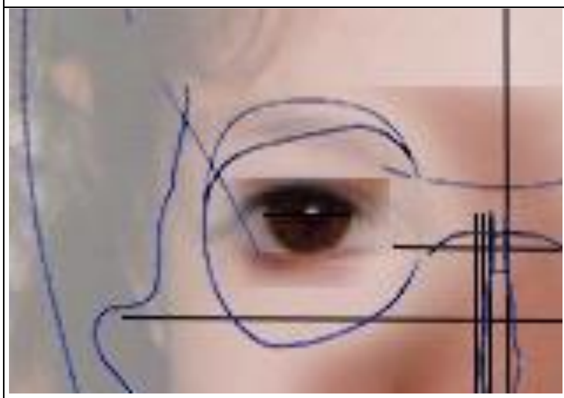
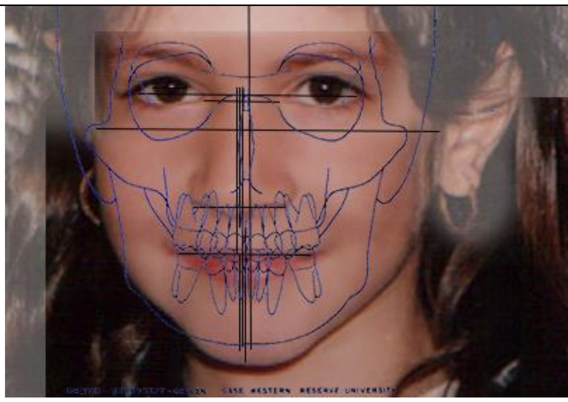
Table 14.2

Female (mm)	zy-zy	ex-ex	en-en	al-al	ch-ch	n-sto	n-sn	n-gn	sn-gn	sto-gn	tr-gn
Age 1	95.6	75.3	26.9	25.9	33.3	46.5	29.2	77.2	47.3	31.4	141.1
Age 2	97.9	75.5	26.6	26.1	35	50.7	32.6	83.8	51.7	34.4	145.8
Age 3	101.2	77.3	27	25.9	36.3	53.4	34.6	86.9	54.3	35.5	148
Age 4	106.8	75.3	29	27.8	37.9	56.1	37.8	92.6	57.8	40.2	145.2
Age 5	109.4	76.5	29.4	28.5	39.5	58	39.3	96.5	59.4	41.3	151.9
Age 6	113.4	77.8	29.8	27.8	41.2	57.9	39.3	95.7	58.8	40.3	155
Age 7	115.8	79.4	30.1	28.6	42.4	59.7	40.7	98.3	59.7	40.7	158.9
Age 8	117.3	79.2	30.5	28.5	43.1	60.4	41.5	98.1	59.3	40.6	159.3
Age 9	119.4	81.4	31.1	29.2	44.6	62.3	43.6	101.3	59.9	40.9	161.6
Age 10	120.7	81.8	31.2	29.6	44.9	63.2	44.5	103.9	62.2	42.5	164.4
Age 11	122.5	82.8	31.6	29.9	45.9	64.4	45.7	104.7	62.1	42.2	164.3
Age 12	123.6	83.6	31.6	30.9	46.5	66.3	47.2	108.2	64.6	44.1	168
Age 13	126.8	85.4	32.2	31	48.1	67.3	48.2	109.1	63.9	43.5	168.6
Age 14	128	85.3	32.4	31	47.5	68.2	49.1	110.7	64.8	44.2	170.8
Age 15	129.7	87	32.7	31.7	49.1	68.8	49.2	111	64.1	43.5	170.7
Age 16	130.6	86.9	31.8	31.6	48.9	70.4	50.4	113.5	65.9	44.7	172.1
Age 17	131.1	87.6	32.5	31.9	49.4	68.9	49.2	112	65.3	44.7	172.7
Age 18	129.9	86.8	31.6	31.4	49.8	68.1	48.9	111.8	65.5	45.2	172.5
Age 19-25	130	87.8	31.8	31.4	50.2	69.4	50.6	111.4	64.3	43.4	173.3

Table 15: Age progression workflow

1: Source image	2: Enhance image	3: closed lips
		
Age 7	Age 7 enhanced	Age 7 lips closed

4: 12 measurements	5: Extrapolate	6: Facial features
		
Age 7 measure in pixels	Extrapolate measurements	Feature placement

7: Progression [manual image manipulation]
 
• Enhance the image to improve the quality before making any changes • ‘Closed lips’ should be depicted before the measurements; an opened jaw should be ‘corrected’ by moving the chin and gonial angle slightly upwards. • Calculate, extrapolated and centre the measurements, Bolton standard can be used as a guideline • Copy and place the facial features (eyes, nose and lips) from the original image (Age 7) onto the extrapolated measurement lines • Stretch the original image (forehead, the width of the face, ears and the lower 1/3 of the face) to match the measurements • The width of the head can be reduced by using the warp or distort tool in “free-transform” • The diameter of the iris and the height of the eyes remain unchanged • Stretch the width of the eyes to match the ex-ex, check the position of en-en • Without changing the size, place the original iris back on top of the stretched eyes • Stretch the width and height of the nose to match the measurements • Stretch the width of the lips to match the measurements

• Add shadows to sides of the nasal body and alter the tip of the nose to be slightly downward pointing (so the nose appears to be less button like, taller and wider) • Widen the chin and jawline according to the estimation, warp or 'liquify' for a more define the jawline • Stretch the dimension of the cheeks from below the eyes if necessary • reposition and deepen the nasolabial folds, and any other creases if necessary • Texture from another individual at a similar age may be used • Darken and thicken the eyebrows if necessary • Blend the features together to generate a final image
Practitioners should make alterations where the depiction remains to be a 'convincing face'. This process helps to guide the 'growth' aspect of the age progression and it is still subjected to the subjectivity with an artistic impression.

8: Final image

Original image Vs Final depiction at age 14 years

The mean difference in iris diameter between 3 months to 8 years old is approximately 0.818mm, with an average size of 10.70 +- 0.73 mm in diameter between all subjects (Ronneburger et al., 2006). This range was between 8.9-12.6mm and showed no significant correlation to the child's age or sex (Ronneburger et al., 2006).

The image was imported to Adobe Photoshop CS6 in Step 1, and enhanced in step 2 (Table 15). Step 3 measured the 12 landmarks including the iris diameter. By adopting the methods proposed by Farkas et al. (1994) and Machado et al. (2017), the image was translated to 'life-size' from pixels (Age 7 est. mm) using the iris diameter as a fixed measurement at 10.7mm (Table 16). The translated measurements in millimetres were compared with the Farkas standard at the original age of the photograph (Age 7). The difference between the target age (Age 14) and source age (Age 7) from the Farkas standard was calculated as a ratio difference (Ratio =Target age/Source age). This ratio was then used to extrapolate the difference in the estimated measurement both in millimetres (age 14 est.mm) and in pixels (est. pixels). Once the measurements were extrapolated, the image was manually manipulated, together with the Bolton standards (Broadbent et al., 1975); a template was used as guidance for the age progression method to produce a final image.

Table 16: Example of measurement estimation

Subject FF073												
Farkas Standard	iris	zy-zy	ex-ex	en-en	al-al	ch-ch	n-sto	n-sn	n-gn	sn-gn	sto-gn	tr-gn
Farkas norm Age 7 (mm)		115.8	79.4	30.1	28.6	42.4	59.7	40.7	98.3	59.7	40.7	158.9
Farkas norm Age 14 (mm)		128	85.3	32.4	31	47.5	68.2	49.1	110.7	64.8	44.2	170.8
Ratio difference		1.11	1.07	1.08	1.08	1.12	1.14	1.21	1.13	1.09	1.09	1.07
Digital image [Iris: Pixel > mm ratio = 0.428]												
Image (pixel)	25	214	144	51	62	81	102	79	168	89	66	307
Age 7 (est. mm)	10.70	91.59	61.63	21.83	26.54	34.67	43.66	33.81	71.90	38.09	28.25	131.40
Age 14 (est. mm)		101.24	66.21	23.50	28.76	38.84	49.87	40.79	80.97	41.35	30.68	141.24
Est. Pixels		236.55	154.70	54.90	67.20	90.74	116.52	95.30	189.19	96.60	71.68	329.99

**est = estimated

By comparing the estimated landmarks to the Farkas Standard, this method resulted in measurements tailored to the image by using a ratio of growth for each landmark. The accuracy of this method will be affected by a number of factors:

- The quality of the source image, such as subject-to-camera distance, definition, lighting, head-pose, facial expression, hair and glasses.
- Individual growth pattern/heritage of the individual (Farkas and Bolton's standards were both based on a North American population)
- Artistic interpretation

Depending on the availability of veridical images within the database, the original image could be used to generate age progression images with a minimum of 2 years age interval at different ages up to 18 years. These original photographs were referred to throughout the text as ‘original’, ‘source’, ‘outdated’ or ‘input’ image/photograph; the manipulated image as ‘progression’ or ‘depiction’; target images as ‘target’, ‘veridical’ or ‘ground-truth’. Because each individual within the database did not have images across every age, age progressions were chosen based on the availability of target images as listed in Table 17 below. Each original image could be age-progressed more than once.

Table 17: FG-NET subjects used for age progression in Experiment 2B

FG-NET: Female			
	Age (years)		
Subject	Source	P1	P2
002	5	12	18
002	7	12	18
008	6	12	17
008	8	12	17
009	3	9	16a,b
009	9	16a	16b
010	4	6	18
010	7	12	18
015	5	12	15
026	2	6	15
026	6		15
049	6	10	15
052	7	14	18
054	6	12	17
059	3	12	16
065	3	7	13
065	7	9	13
072	7	13	17
073	4	9	14
073	7	12	14
076	6	10	16
076	8	12	16

FG-NET: Male			
	Age (Years)		
Subject	Source	P1	P2
011	5	13	17
011	13		17
035	7	14	18
035	9	16	18
037	6	13	17
037	8	13	17
044	5	7	17
046	4	10	17
057	5	16	18
058	5	10	17
063	2	18	
066	2	7	11
066	4	9	11
068	3	10a,b	14
074	6	10	15
074	8	13	15
075	2	8	11
075	6	8	11
081	3	8	12
081	5	8	12

P1 = First progression
P2 = Second progression

FG-NET = F
Female = F
Male = M

FG-NET female subject 002 = FF002
FG-NET Male subject 011 = FM011

FG-NET: n= 27 (14F,13M)
Original images: 42 (22F, 20M)
Progressions: n=80 (42F, 38M)
Matches: n= 83 (44F, 39M)

Based on the 27 FG-NET subjects (14F, 13M), 80 progressions were made:

- 42 original images (22F, 20M) were selected to generate 80 age progressions
- 39 original images (21F, 18M) had two progressions P1 and P2, and three (1F, 2M) with one age progression
- 15 subjects (8F, 7M) had two different source images to generate 57 depictions (30F, 27M) of various ages
- Of those 15 subjects, 42 depictions were generated to the same age (20F, 22M) from two different original images
- A total of 83 comparisons were made, as three progressions (2F, 1M) had two target images of the same age for comparison (i.e. subject FF009, and FM068)

4.2 Experiment 2A: Inter-observer study

Using a subject from the OACAD database, one original image at 3 years old from subject OAM002 was chosen (Figure 32). Three practitioners were asked to generate two age progressions to age 6 and 13 years old based on the original image using the guided method described in 4.1.1.

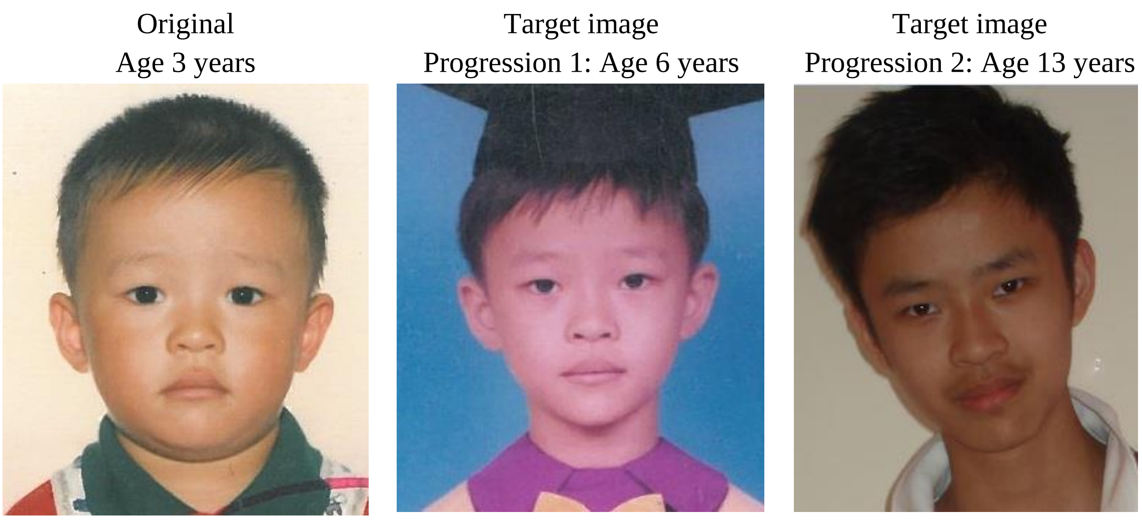


Figure 32: An example of manual age progression using guided methodology (OAM002) from the inter-observer test

1. Measurements taken by the practitioners were compared
2. The likeness between the progressions was compared manually
3. The age progressions were compared to the target images using Microsoft Face API

4.3 Experiment 2A: Results

Two age progressions were generated by three practitioners, and four practitioners participated in the measurement of the original image.

4.3.1 Measurements comparison

Measurements using the guided method were compared between four practitioners. A percentage difference was based on the mean and Table 18 shows that measurements involving the landmark Nasion (n) and Trichion (tr) were the most variable between 13-37%. This suggests nasion could be an inconsistent landmark. The hairline of the subject was often masked, which would have led to the inaccurate approximation of the landmark Trichion. Other measurements were consistent between the practitioners with a difference between 0-8% when compared to the mean.

Table 18: Inter-observer measurement statistics in pixels

Landmarks	Min	Max	Q_1	Median	Q_3	Mean	difference	difference%
sto-gn	52	52	52	52	52	52	0	0
iris	17	17.06	17	17	17.015	17.015	0.06	0.35
zy-zy	186	188	186.75	187.5	188	187.25	2	1.07
ex-ex	113	115	113	114	115	114	2	1.75
ch-ch	55	56	55.75	56	56	55.75	1	1.79
en-en	50	51	50.75	51	51	50.75	1	1.97
sn-gn	78	82	79.5	80.5	81.25	80.25	4	4.98
al-al	46	50	46.75	47.5	48.5	47.75	4	8.38
n-gn	132	151	135	136	139.75	138.75	19	13.69
tr-gn	203	236	223.25	230.5	232.25	225	33	14.67
n-sto	80	100	83	85	89.5	87.5	20	22.86
n-sn	51	73	53.25	56.5	62.5	59.25	22	37.13

4.3.2 Manual comparison

The outline of the progressions from age 3 to 6 years old (Figure 33) and 3 to 13 years old (Figure 34) were compared to the target images. The best fit was based on the proportion of the internal features, i.e. positioning of the eyes, nose and mouth.

A	B	C
		

Figure 33: Inter-observer manual comparison of age progression image outlines with the target image (progression from age 3 to 6 years old)

A	B	C

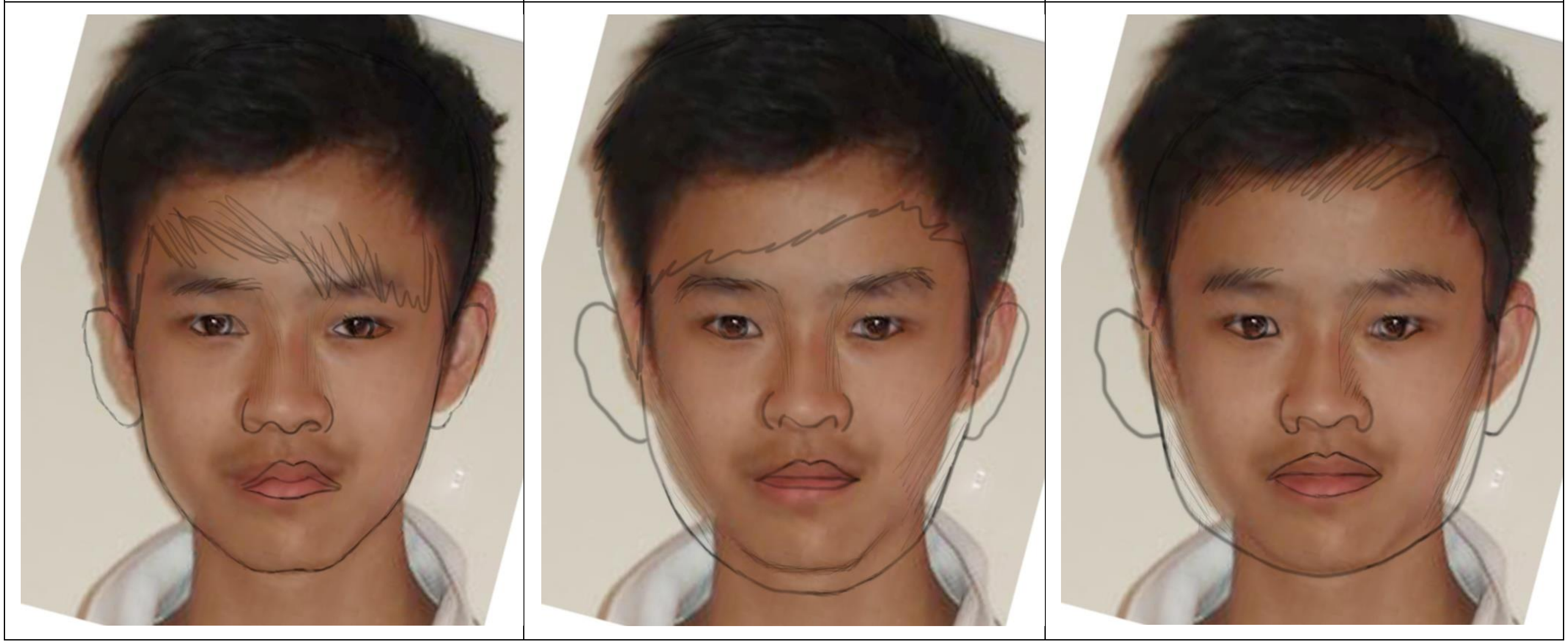


Figure 34: Inter-observer manual comparison of age progression image outlines and the target image (progression from age 3 to 13 years old)

In relation to face height, the positioning and proportion of the facial features of the progressions in Figure 33 and Figure 34 varied between practitioners, especially to the lower 1/3 of the face in Figure 33. This could be caused by the inconsistent approximation of the landmark Nasion and Trichion. Facial features that varied the most between practitioners were: face shape, jawline, lip shape, and the positioning of the eyebrows. These features were not guided by the measurements and therefore varied with artistic interpretation. The most consistent feature was the prediction of nose width and shape in relation to the rest of the face.

4.3.3 Comparison using Microsoft Face API

Figure 35: Inter-observer comparison images (Microsoft Face API)

In comparison to the target image at age 6, although the manual comparison in Figure 33 suggests depiction A was more dissimilar in the lower third of the face, this depiction created by Practitioner A yielded the highest confidence score using Microsoft Face API (Figure 35). However, the face shape of depiction A was more similar to the target when compared to B and C. Although the positioning of the facial features in depiction B was very similar to the target image, the shape of the individual features was more dissimilar when compared to A and C, this could have resulted in the lower confidence score.

Depictions for age 6 and 13 created by practitioner A and C both achieved a higher confidence score in comparison to the original images. This suggests the performance of an age progression varies between practitioners, where some could generate a higher confidence score.

Results above suggests the measurements are able to guide the positioning of the features to a certain level, however, the process of age progression remains to be variable with artistic interpretation.

4.4 Experiment 2B: Application of conditions for machine-based face recognition

Computer scientists have explored how image quality can affect facial recognition systems (FRS) (Dodge and Karam, 2016; Grm et al., 2017), but when dealing with age progression images, researchers have suggested that inaccurate information can often have a negative effect on human recognition (Claes et al., 2010a, 2010b; Mahoney and Wilkinson, 2012). Psychologists have tested how different conditions can affect face recognition, conditions such as colour, illumination, low resolution (Sinha et al., 2006). Gaussian blurring in face recognition suggested that faces are recognisable even when they are blurred or pixelated (Bachmann, 1991; Hole et al., 2002; Lander et al., 2001). External features such as hair can also have an effect on human recognition (Erickson et al., 2016; Frowd et al., 2012; Toseeb et al., 2012). With the conditions above, Experiment 2B aims to explore how an FRS is able to handle 'resemblance' images at different conditions known to benefit human recognition.

To investigate how different image manipulations can affect the recognition score using Microsoft face API, the age progression images (Table 17) were subjected to different conditions using photo-editing software (Adobe Photoshop CS6):

1. Condition 1: Does the recognition rate improve by removing colour from the image?
2. Condition 2: Does the recognition rate improve by cropping the image?
3. Condition 3: Does the recognition rate improve by reducing the image resolution?
4. Condition 4: Does the recognition rate improve by blurring the image?
5. Condition 5: Does the recognition rate improve when the best conditions are combined?

[Example subject: Subject FM035 age 16 years progression from age 7 years, conditions shown in Figure 36 to 39]

Figure 36: Image condition 1: Black and White

Colour	Black and White
	

Some original images were black and white; these datasets were excluded from the comparison of condition 1.

Figure 37: Image condition 2: Cropped

Colour	Black and White
	

Some original images did not show the hairline and were excluded from the comparison of condition 2.

Figure 38: Image condition 3: Resolution reduction

Original (441X568)	20% (353X455)	50% (265X342)	70% (176X227)
			

The image quality and size varied and therefore the image resolution was not the same. Rather than standardising all images to the same pixels, the original pixels of the image was documented and be reduced to 20%, 50% and 70% (Figure 38). The setting with a higher confidence score represented the performance of this condition. It should be noted that the resolution was reduced via the image size and not pixelated using the Photoshop filter.

Figure 39: Image condition 4: Gaussian Blur

Original	1 pixel	3 pixels	5 pixels
			

Similar to the resolution above, the quality of the images varied. Therefore, the setting with a higher confidence score represented the performance of this condition. The Gaussian blur filter was applied as a function in Adobe Photoshop CS6. The images were blurred with 1 pixels, 3 pixels and 5 pixels.

Image condition 5: Combined

Based on the recognition scores of the tests above, the best conditions with a higher confidence score were combined, for example, 70% resolution reduction and cropped.

4.4.1 Comparisons

The original (out-dated) image and the manual age progression images were compared to the 'veridical' image (image of the individual at the target age) using Microsoft Face API. This particular algorithm gives a confidence score between two facial images. With two decimal places, a score above 0.5 displayed 'two faces belong to the same person' and a score below 0.5 displayed 'two faces not belong to same person' (Figure 40). This would provide an objective assessment of the 'likeness' between two images.

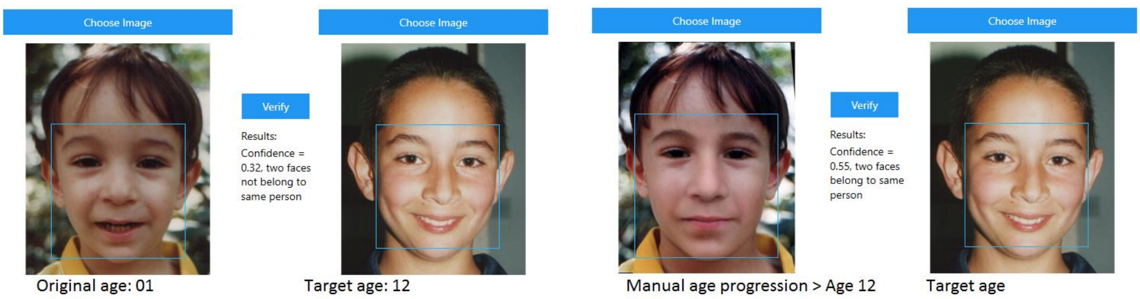


Figure 40: Microsoft Face API confidence scores

The confidence score generated by Microsoft Face API between the manual age progressions and the target image was compared to the score between the original and the target image; all comparisons made were documented in the Appendix. Age progression was compared with different conditions, and image variability was assessed using the three cases with more than two target images, from subjects FF009 and FM069.

<u>Limitations</u>:

- Quality of the images differs
- Only available as single source comparison
- The method used in manual age progression differs between practitioners, therefore the depictions are subjective
- The target age for each progression is dependent on the available 'veridical' images within the data for each individual

4.4.2 <u>Manual facial comparison</u>

When the original image performed better than the depiction, this suggests that lower recognition of the depiction was unlikely to be due to image quality. To evaluate the age progression and provide suggestions for improvements, cases with a score difference above 0.9 between the images (age progression and original) and the target were selected for manual facial comparison in 4.5.4. The image set with the lowest confident score was also compared.

The facial comparison is a task performed by humans for intelligence gathering, identity management, screening/access control, investigative/operational tool or forensic identification (FISWG, 2010). Following the guidelines provided by the Facial Identification Scientific Working Group (FISWG), the four main methods were holistic comparison, morphological analysis, photo-anthropometry and superimposition (FISWG, 2012). FISWG only recommends morphological comparison for forensic investigation.

In a real case scenario it is unlikely that a target/veridical image will be taken from the same viewpoint as the original. Since photo-anthropometry is sensitive to image quality (FISWG, 2012), anthropometric comparison was not advised as a method of facial comparison by

FISWG (2012), as it can significantly reduce the accuracy of the analysis. However, due to the methodology set up of the age progression method, the proportional anthropometric extrapolation from the age progression was compared to the target image. With a change in head pose, the positions of the horizontal measurements will not be comparable. A change in the horizontal position of the measurements will not affect the vertical proportions, measurements i.e. iris diameter, face width (zy-zy), nasal width (al-al), mouth width (ch-ch) were moved horizontally to fit a veridical image of a different head pose.

Superimposition can often mask and blur the details of two images making comparison difficult. Therefore the feature outlines of the veridical image were traced using Adobe Photoshop CS6. The outline was superimposed onto the age progression and scaled using the iris diameter.

Nineteen facial components were considered as a standard procedure of facial comparison analysis recommended by FISWG (2013):

1 Skin; 2 Face/Head Outline; 3 Face/Head Composition; 4 Hairline/Baldness Pattern; 5 Forehead; 6 Eyebrows; 7 Eyes; 8 Cheeks; 9 Nose; 10 Ears; 11 Mouth; 12 Chin/Jawline; 13 Neck; 14 Facial Hair; 15 Facial Lines; 16 Scars; 17 Facial Marks; 18 Alterations; 19 Other

These components were analysed individually for each case. It is worth noting that human comparison will differ to how an FRS recognises a face. The differences or similarities between each facial component were described according to the FISWG guidelines. Utilising the information from the guideline, Table 19 is tailored to the comparison in this study analysing near frontal images.

Table 19: Facial Components comparison (adapted from FISWG (2013))

1	Skin	Overall texture and tone (Luminance and colour)
2	Face/Head Outline	The shape of the cranial vault and face (Portrait/profile contour description)
3	Face/Head Composition	Proportions/position of features (compare the predicted age progression measurements to the target image)
4	Hairline pattern	The shape of the hairline
5	Forehead	Relative height and width, brow ridge prominence continuity
6	Eyebrows	Asymmetry; shape; size; tilt (in relation to the medial and lateral canthus); hair details, density and distribution.
7	Eyes	Inter-eye distance; shape and angle; eyelid prominence, protrusion, visibility of the eye creases/folds; position in relation to the iris; eyelash characteristics; eyeball prominence; sclera colour and blood vessels; iris colour, visibility, diameter, position, irregularity; shape and angle of the medial canthus, caruncle; shape and angle of the lateral canthus; asymmetry in shape and angle.
8	Cheeks	Cheekbone prominence; dimples
9	Nose	Overall shape, length, width, prominence and symmetry; nasal bridge width, length, shape and depth; nasal body width length shape and angle; nasal tip shape, angle and symmetry; nasal base width, height and deviation; alae thickness, symmetry and shape; nostrils shape, size, symmetry and hair; columella width, length, relative position, symmetry.
10	Ears	Asymmetry, size, shape, protrusion and positioning.
11	Mouth	Philtrum prominence, width of ridges and furrow, symmetry; overall mouth shape and symmetry; upper and lower lip (shape, fullness, protrusion symmetry, vermilion border shape, details); lip fissure shape, symmetry, degree of contact and corners; asymmetry; dental occlusion; prognathism; dentition shape, size, alignment, condition; abnormalities.

12	Chin/Jawline	Chin shape, relative length, prominence, symmetry, details; jawline shape and definition; gonial angle shape and definition.
13	Neck	(not aged for FRS, therefore not comparable)
14	Facial Hair (male)	Shape distribution, texture, symmetry, density, variation in colour, orientation, edge definition, continuity, around the mouth, long hairs.
15	Facial Lines	Forehead, nasion, crow’s feet, infraorbital, lip creases, nasolabial folds (distribution, orientation, quantity, pattern, depth); upper and lower eyelids visibility, position, depth and shape; marionette lines and cleft chin.
16	Scars	Location, shape, orientation, size, colour, depth
17	Facial Marks	Freckles, moles, acne, rosacea, birthmarks, bruises, abrasions, vitiligo, dark/light patches: location, shape, size, colour, prominence
18	Alterations	Piercing, makeup, tattoo and other (location and description)
19	Others	Irregular feature description

4.5 Experiment 2B: Results

Based on the 24 FG-NET subjects (14F, 13M), 42 original images (22F, 20M) were selected, with most having 2 progressions to different ages, 80 age progressions were generated using the guidance from Experiment 2A, and 83 comparisons were generated.

4.5.1 Original and Age progression versus target

The original image, age progressions, and progressions with conditions were all compared to the veridical/target image using Microsoft Face API. To carry out a chi-square analysis, the conditions (i.e. black and white, cropped, resolution reduction, blurred and combined) were collated as ‘Age progression’ (Figure 41). For each individual, the image type with the highest confidence score was recorded. If the score was the same between the image types, both were recorded as the ‘highest’.

58% (52/90) of the age progression depictions performed better than the original image. However, the chi-square test found no statistically significant difference between the recognition rates for the original image and the age progressions [$\chi^2(1) = 2.178$, $p = 0.140$], and no statistically significant difference between the recognition rates for the male and female subjects [$\chi^2(1, N = 146) = 0.247$, $p = 0.619$]. Appendix 2 details the statistics report.

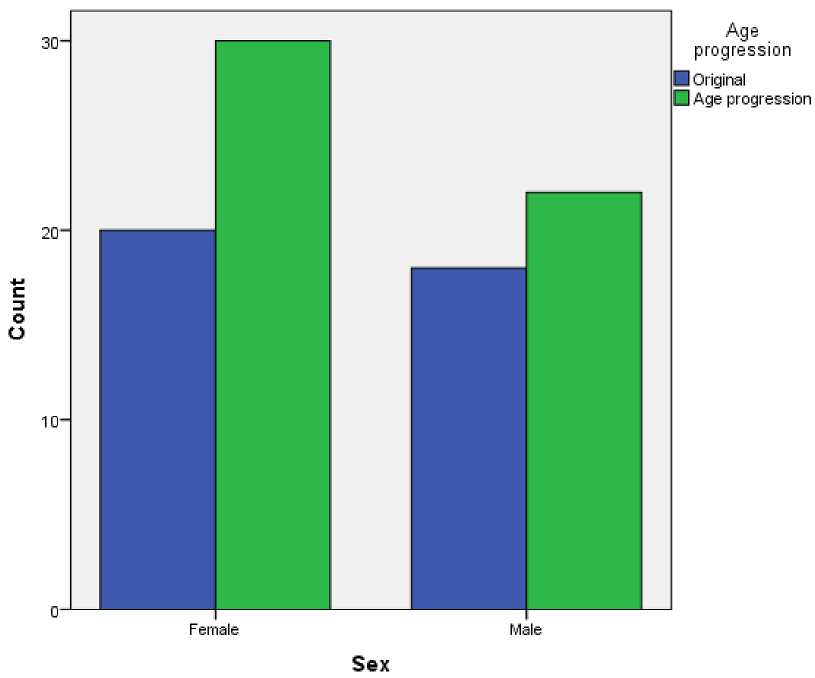


Figure 41: Highest confidence score count for original images and age progression images

When the veridical/target images were compared to the progression with the highest confidence score, 69% (57/83; 30F, 27M) were recognised by Microsoft Face API as the same individual (confidence score >=0.5), and 31% (26/83; 14F, 12M) failed to be recognised as the same individual (Confidence score <0.5).

A between-subjects univariate analysis was conducted to compare the effect of the age gap on subject sex and image type. This analysis supports the Chi-square test, suggesting no significant difference between subject sex [F(1,120)=0.764, p=0.384], and image type [F(1,120)=0.187, p=0.666]. Figure 42 suggest that as age gap increases, recognition decreases, and this interaction is significant [F(13,120)=15.765, p<0.000]. Appendix 2 details the statistics report.

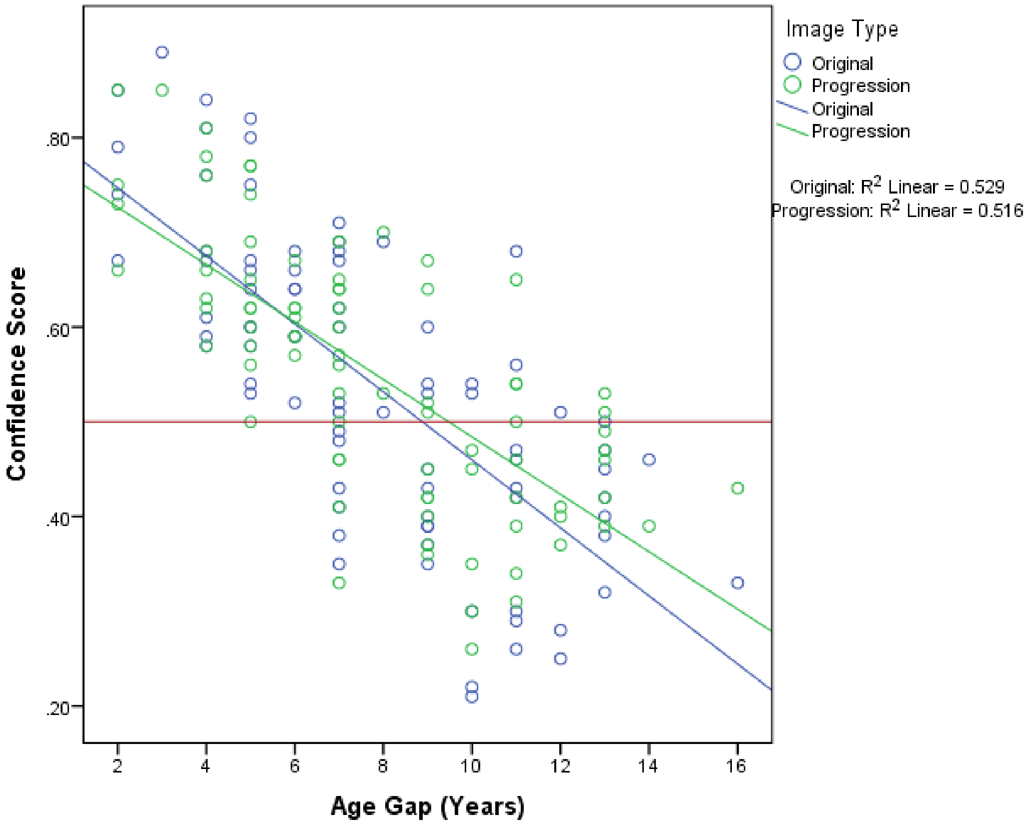


Figure 42: Face recognition rate for age gap (Years) related to the image type (Original and Progression) versus the target image (highest confidence score count)

Figure 42 combines the male and female data showing the highest confidence score between original and age progressions. The line of best fit suggests this Face API fails to recognise the individual at an age gap of approximately 9 years.

A between-subjects univariate analysis was conducted to compare the effect of original age on recognition when the data was split by subject sex and image type. Although the test suggests no significant effect between subject sex, image type, and original age [F(7,132)=0.096, p=0.998], there was a significant interaction between sex and original age [F(7, 132)=2.355, p=0.027]. However, by combining the progression types showing the highest confidence score between male and female data, there was no significant correlation between confidence score with original age [r=0.005, n=166, p=0.947]. Appendix 2 details the statistics report.

Figure 43 suggests a positive trend between the confidence score and original age for female subjects and a negative trend for male subjects, however, these trends were not statistically significant.

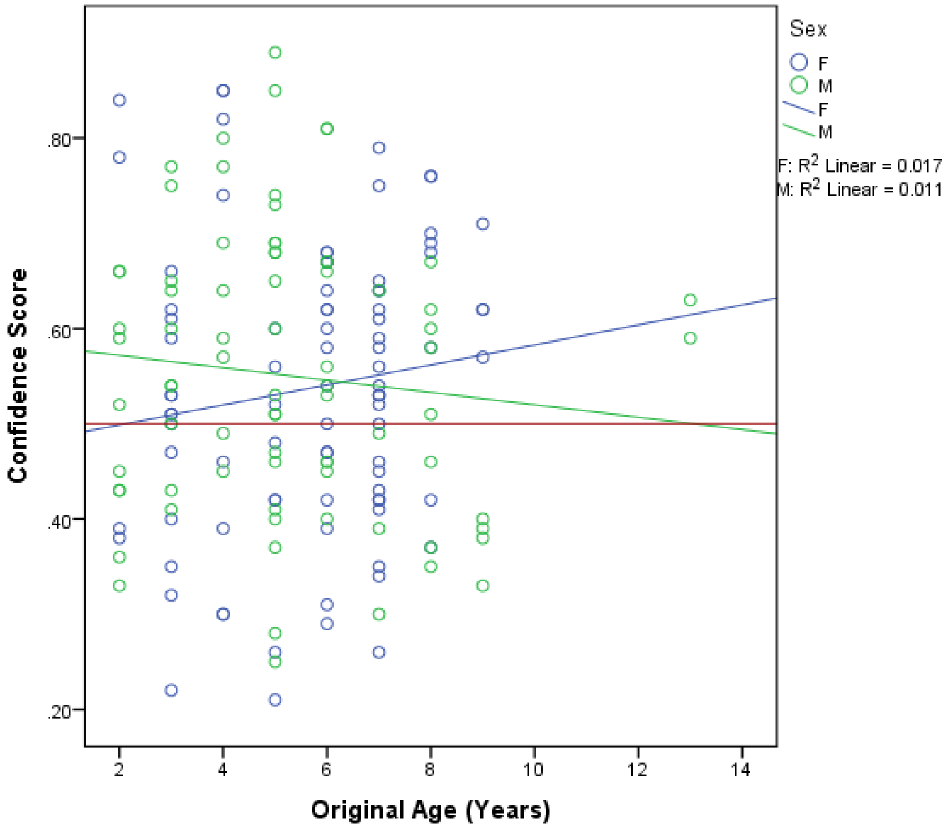


Figure 43: Face Recognition rate for original age (Years) related to sex (Male and Female) versus the target image (highest confidence score count)

Could the original age influence the correlation between the age gap and confidence score? When the original age was controlled, partial correlation shows that there was a significant correlation between the age gap and confidence score [r(163)=-0.741, n=166, p <0.001].

However, without controlling the original age, the statistical significance remained very similar [$r(164)=-0.720$, $n=164$, $p<0.001$]. This suggests original age had very little influence on the interaction between the age gap and confidence score. Appendix 2 details the statistics report.

4.5.2 Age progression conditions

Each age progression was manipulated with conditions black and white, cropped, blur, resolution reduction and combined. Condition(s) with the highest confidence score were recorded; if the score was the same between multiple conditions, all conditions were recorded as the 'highest'; if scores were the same in comparison to the age progression, it was marked 'None'. Chi-square tests were performed and found a significant difference between the conditions [$\chi^2(6, N=123)=50.520$, $p<0.001$], with no significant difference between the subject sex [$\chi^2(1, N=123)=0.984$, $p=0.321$]. Appendix 2 details the statistics report.

This suggests that although the depictions remained the same, recognition in Microsoft Face API can be affected by a change in condition, especially resolution, blur and combined.

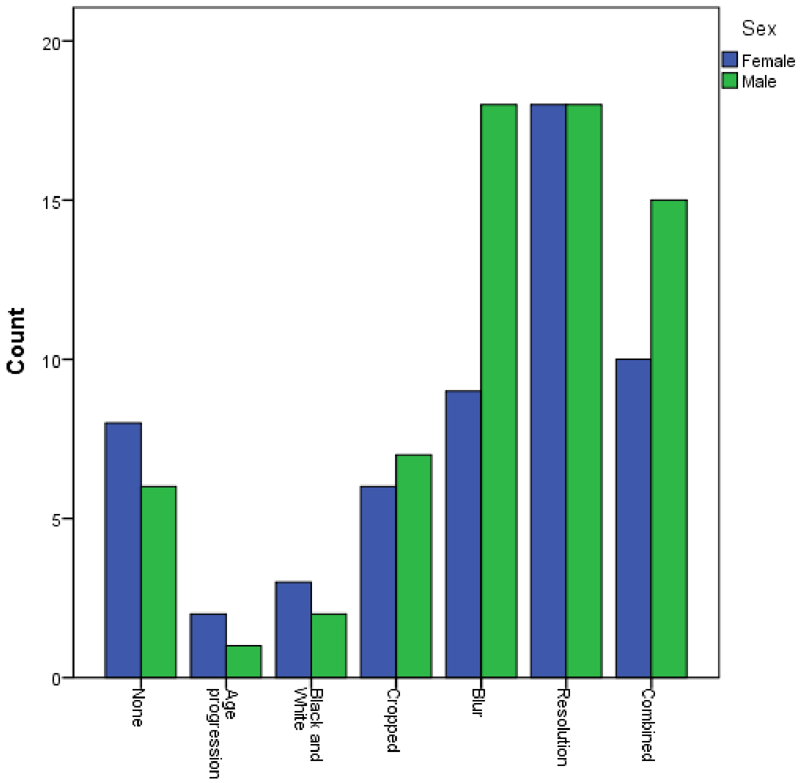


Figure 44: Highest confidence score count for the different age progression conditions when compared to the target image

To explore if a condition reduced recognition, for each individual, confidence scores of the conditions were compared to the age progression. Conditions with the lowest score were documented; if scores were the same, higher or with no lower scores in comparison to the age progression, it was marked 'None'. Chi-square tests were performed and found a significant difference between the conditions [$\chi^2(5, N=147)=58.755, p<0.001$], with no significant difference between the subject sex [$\chi^2(1, N=147)=1.150, p=0.284$]. Appendix 2 details the statistics report. Figure 45 suggests different conditions can have a negative effect on recognition, especially the black and white and cropped condition.

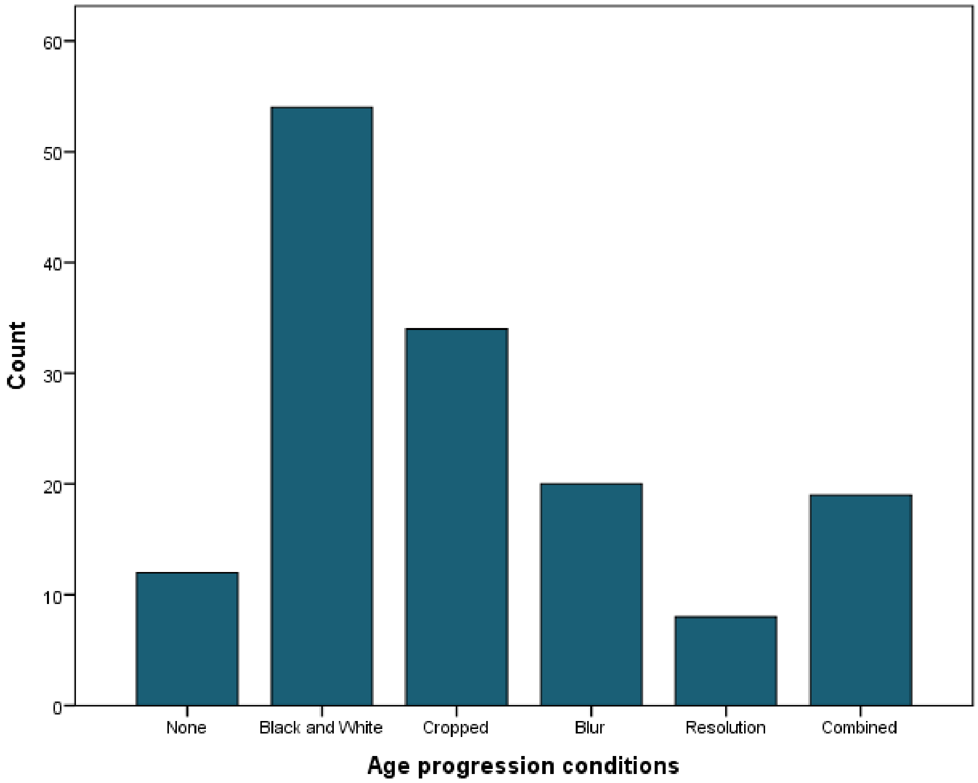


Figure 45: Lowest confidence score count for the different age progression conditions when compared to the target image

4.5.3 Image variability

Most comparisons only had one target/veridical image for comparison with the exception of three comparisons from FF009 and FM069, where more than one target image at the same age was available (Figure 46).

Figure 46.1: FF009 Age 03 > Target age 16A and 16B

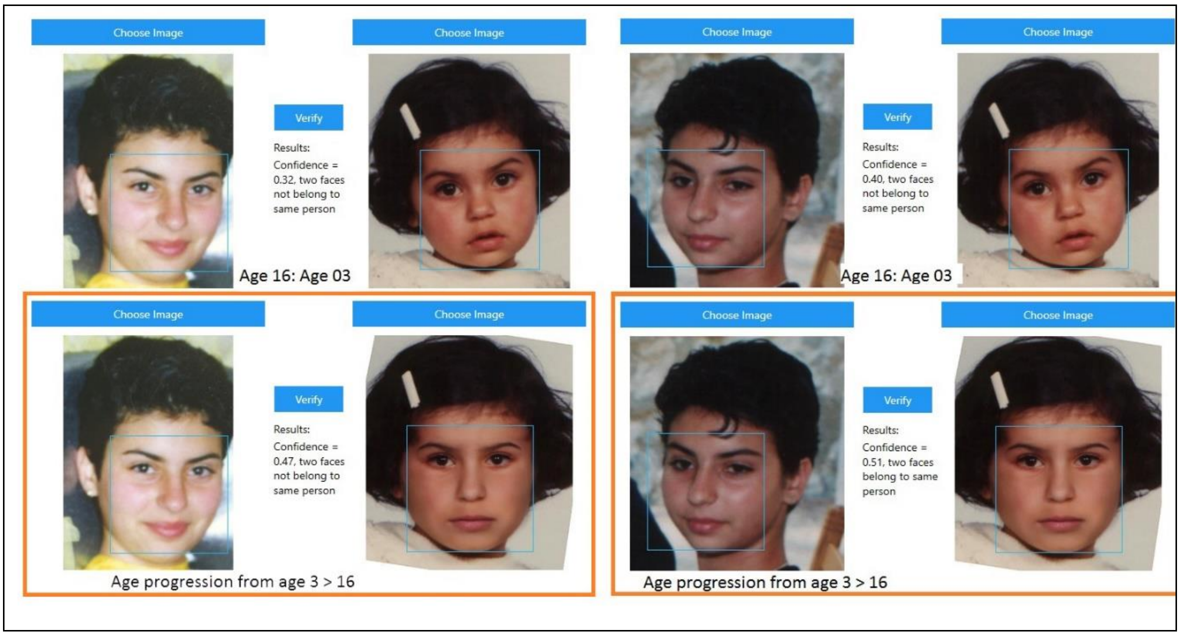


Figure 46.2: FF009 Age 09 > Target age 16A and 16B

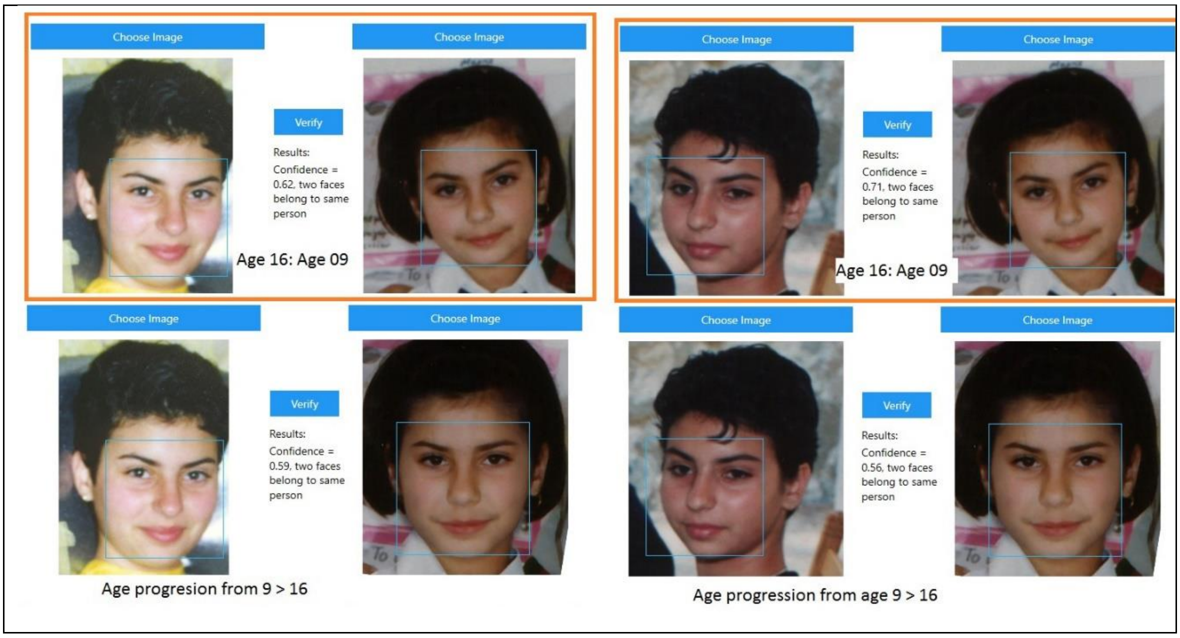


Figure 46.3: FM068 Age 03 > Target age 10A and 10B

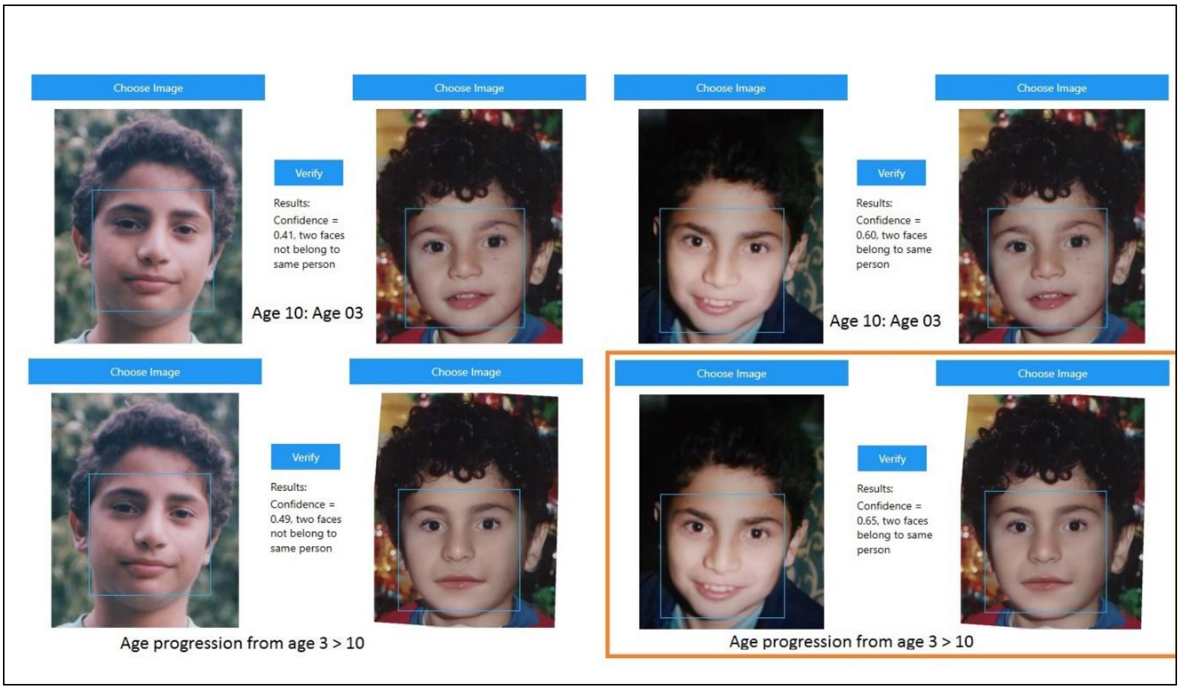


Figure 46: Image variability of different images of the same age when compared to an original image and an age progression

Figure 46 showed the difference in recognition rate when an original image or an age progression was compared to different target images of the same age. In some cases (Age progression in Figure 46.1 and Figure 46.3), a different target image has led to a 'positive' (>0.5) match. This suggests that the difference in image quality can have an effect on the comparisons and could potentially lead to an identification.

To have more than one target image for comparison would be ideal, this information could potentially be available in the process of identification of indecent images of children within a database. However, in a research setting using databases such as FG-NET, the availability of images is limited.

When compared to the target, the difference between original images and age progression was compared to the age gap (Figure 47), there was no correlation between the two variables [r=-0.049, n=83, p=0.663]. Appendix 2 details the statistics report. A negative score difference would indicate the age progression performed better, whereas positive indicates original performed better. Results with no correlation were somewhat surprising, as age progression was assumed more dissimilar to the original image with an increase in the age gap. This suggests the dissimilarities between the original and the age progressions did not increase with the increasing age gap.

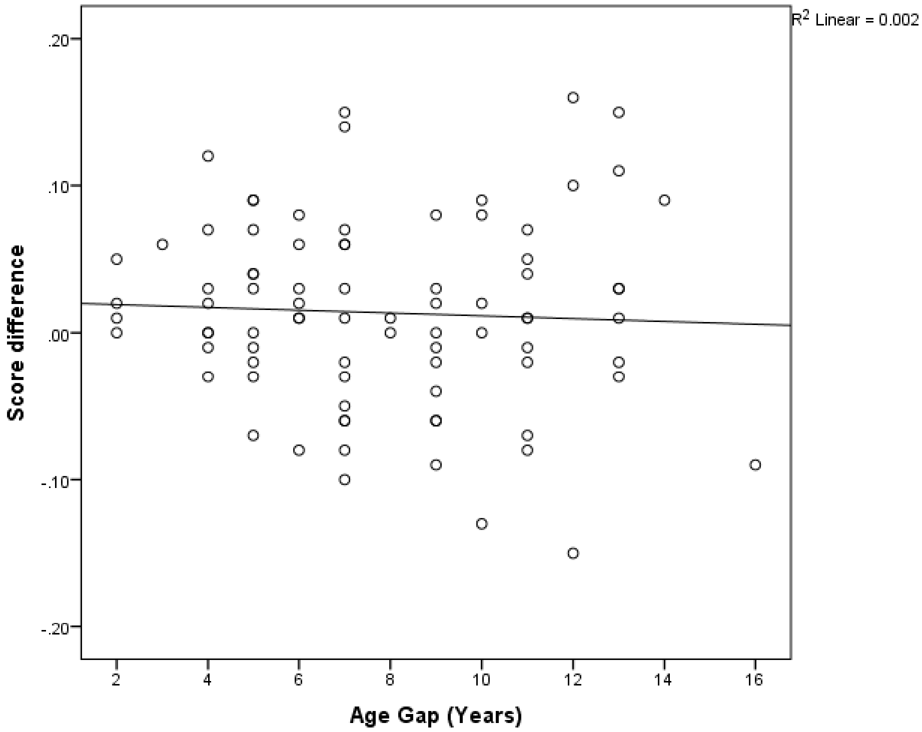


Figure 47: The confidence score difference between the original images and the age progression images in relation to age gap (Years)

4.5.4 **Manual comparison of selected cases**

Four cases (2F, 2M) with a discrepancy scores above 1.0 along with a case (1F) showing the lowest confidence score were selected (Table 20).

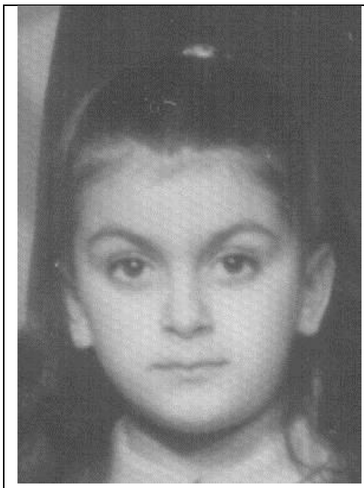
Table 20: Age progression cases with a score difference above 0.9 (compared to the original)

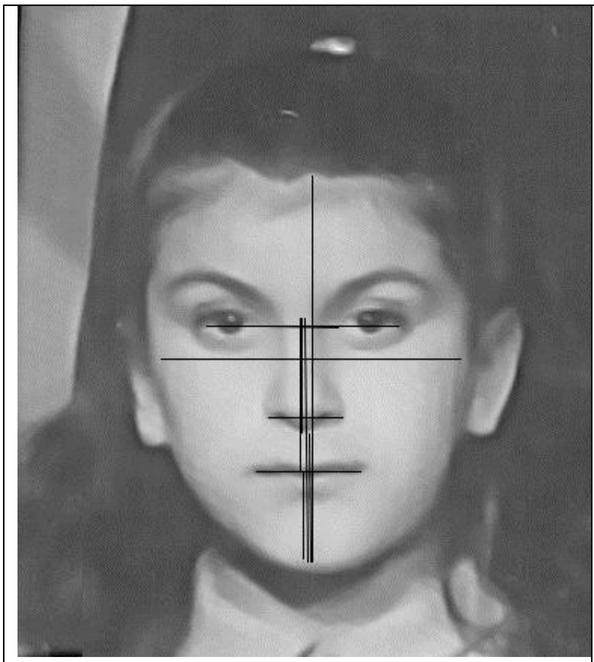
Subject	Age of progression from and to (years)	Age gap	The confidence score of age progression	Score difference
FF008	8 > 12	4	0.56	0.12
FF009	9 > 16a	7	0.51	0.15
FM037	6 > 13	7	0.53	0.14
FM044	5 > 17	12	0.35	0.16
FF015	5 > 15	10	0.21	0

4.5.4.1 **FF008 comparison**

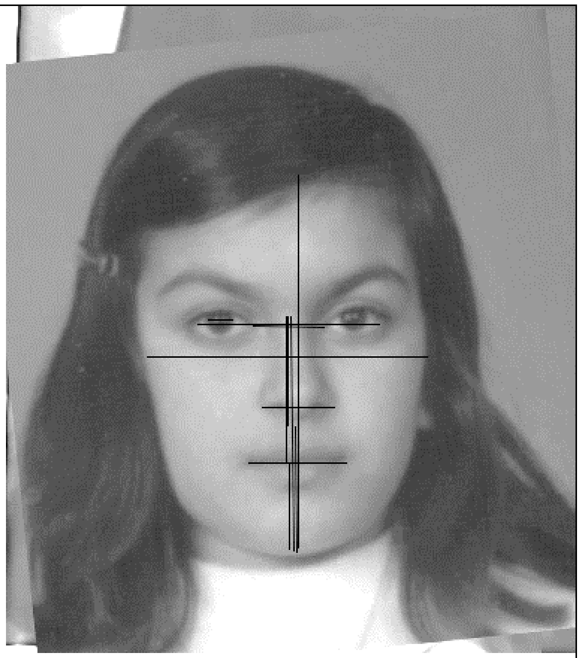
Table 21 compared the age progression of FF008 from age 8 years old to the target image at 12 years old. The age progression and the target image were both frontal views. The left side of the face was shown more in the age progression and the right side of the face was shown more in the target image. The age progression was black and white, the same as the source image; the target image was a ‘red-tinted’ colour faded photograph. Table 22 showed the morphological facial components comparison.

Table 21: FF008 Age 8 > 12 comparison images

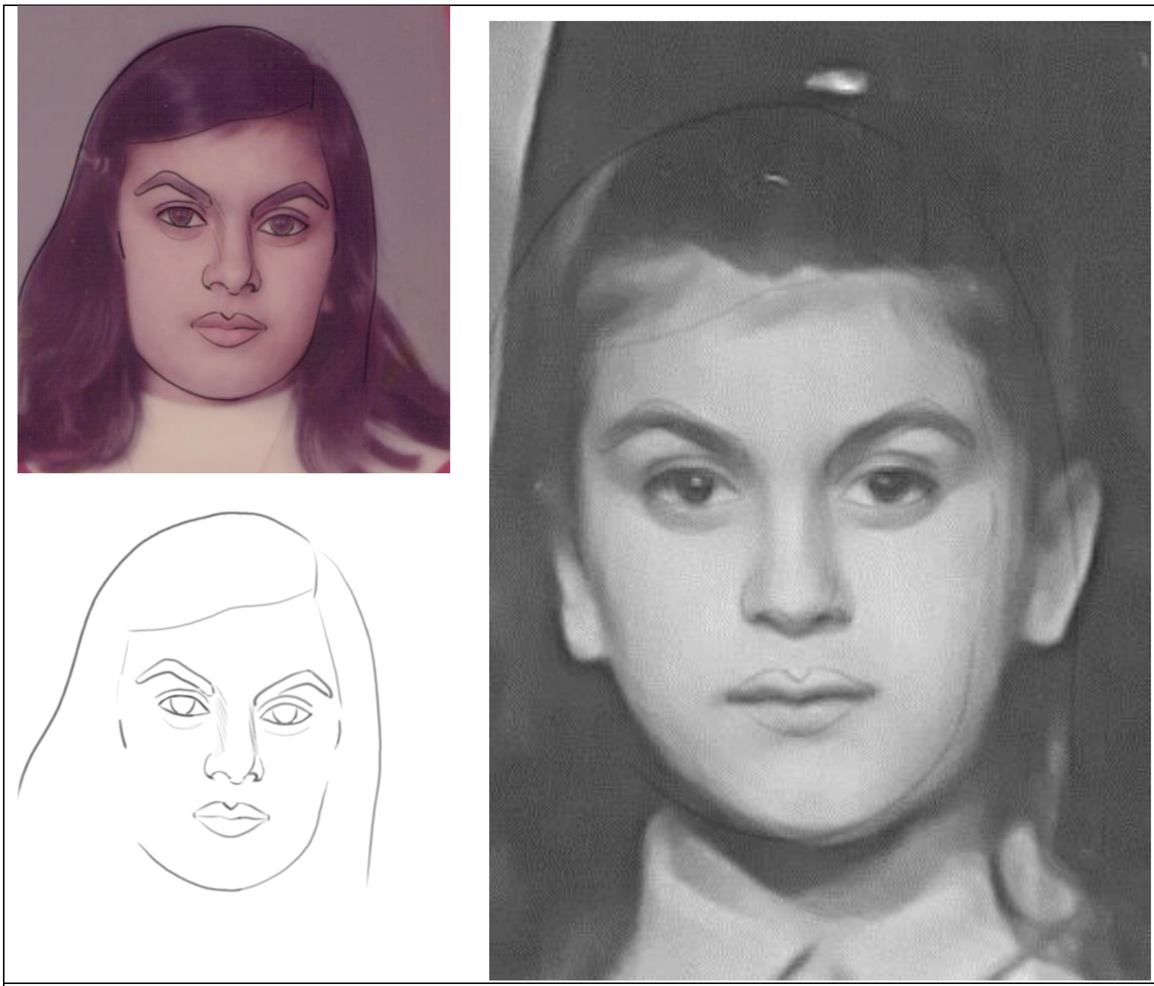
		
Original age 8 years	Age progression to age 12 years	Target/veridical image at age 12 years



Prediction measurements



Comparing prediction to the target image



Superimposing feature outlines of the target/veridical image onto the age progression

Table 22: FF008 age 8 > 12 Facial Components comparison

1	Skin	After the original image was enhanced, the age progression appeared to be more contrasted, thus more luminous.
2	Face/Head Outline	The shape of the cranial vault was similar, but the lower 1/3 of the face appeared to be more squared in the veridical.
3	Face/Head Composition	The proportions/position of features were similar when compared to the predicted age progression measurements.
4	Hairline pattern	The hairstyle was different, but the shape and position of the hairline were similar
5	Forehead	Relative height and width were similar
6	Eyebrows	Eyebrows were similar in position but more tilted and angular in the veridical. The hair density and distribution were similar between the age progression and the veridical.
7	Eyes	The superimposition and the predicted measurements (en-en) suggests that the iris diameter in relation to the position of the inter-eye distance was similar. The inner corners of the eyes were similar in shape and position; the outer corners were wider and more upturned in the veridical. The inferior and superior palpebral furrows (creases) were similar in length and position. The eyelashes appeared longer and denser in the veridical. The eyeball prominence was similar, but more sclera was visible on the lateral ends in the veridical. The shape and angle of the medial canthus were similar, but it was not well defined in the depiction. No obvious asymmetry observed.
8	Cheeks	The positioning of the cheekbones was similar
9	Nose	In relation to other facial components, the shape, length and width of the nose were similar; the tip of the nose was more prominent in the veridical. The nasal bridge was similar in width but more rounded in the veridical. The nasal body was similar in width and depth, but more rounded in the veridical. The nasal tip was similar in shape and angle but was more prominent in the veridical. The nasal base was similar in width and height. Alae shape and thickness were similar, but the nostrils were more visible in the veridical.
10	Ears	Assessment not available
11	Mouth	Due to the quality of the original image, the comparison of the philtrum was difficult, but the proportion and the position were similar. The lip thickness of the veridical was fuller and thicker. The shape of the Cupid’s bow was more defined, and the lip fissure was also straighter in the veridical. The protrusion and the characteristic of the slightly upturned corners of the mouth were similar, but the maxilla was more protruded in the veridical.

12	Chin/Jawline	Chin shape in both images was rounded with similar relative length and prominence, however, the jaw was wider and more angular in the veridical thus giving a wider and more defined gonial angle.
13	Neck	(not aged for FRS, therefore not comparable)
14	Facial Hair	(This is a female subject)
15	Facial Lines	The mentolabial sulcus, lower lid creases and the infraorbital creases were similar in position and prominence.
16	Scars	No noticeable facial scars observed
17	Facial Marks	No obvious facial markings observed
18	Alterations	No facial alterations observed
19	Others	N/A

4.5.4.2 FF009 comparison

Table 23 compared the age progression of FF009 from age 9 years old to the target image at 16 years old. The age progression was near frontal view, but the head pose of the target image was angled at about three-quarters, and the gaze of the subject was not directed at the camera. The left side of the face was shown more in the target image. Table 24 showed the morphological facial components comparison.

Table 23: FF009 Age 9 > 16 comparison images

		
Original age 9 years	Age progresion to age 16 years	Target/veridical image at age 16 years

<table>
<tr><td></td><td></td></tr>
<tr><td>Prediction measurements</td><td>Comparing prediction to the target image</td></tr>
<tr><td colspan="2">  </td></tr>
<tr><td colspan="2">Superimposing feature outlines of the target/veridical image onto the age progression</td></tr>
</table>

Table 24: FF009 Facial Components comparison

1	Skin	The overall texture and tone were darker with more redness in the veridical.
2	Face/Head Outline	Both images had a rounded head and face shape, but the head was bigger in the age progression with a difference in hairstyle.
3	Face/Head Composition	The proportions/position of features were similar when compared to the predicted age progression measurements.
4	Hairline pattern	The shape of the hairline was lower in the veridical.
5	Forehead	The forehead was shorted in the veridical.
6	Eyebrows	The eyebrows were similar in shape and position, but the right lateral end was more downward tilting towards to the lateral canthus; possibly affected by the head pose. The eyebrows were thicker in the veridical.
7	Eyes	The superimposition and the predicted measurements (en-en) suggested that the iris diameter in relation to the position of the inter-eye distance were similar. The general shape and angle were similar, but when the image was superimposed, the left eye of the veridical was wider; possibly affected by the head pose. The inferior and superior palpebral furrows (creases) were similar in length and position. The eyeball prominence was similar but the iris in the age progression was more visible with a whiter and more defined sclera. The shape and angle of the medial canthus were similar. No obvious asymmetry observed.
8	Cheeks	Cheekbones position were similar but more prominent in the veridical.
9	Nose	In relation to other facial components, the shape, length and width of the nose were similar; the tip of the nose was more prominent in the veridical. The nasal bridge of the veridical was narrower in width and the nasal body was straighter. The nasal tip was similar in shape and angle but was more prominent in the veridical. The nasal base was similar in width and height. Alae shape, visibility and thickness were similar.
10	Ears	The left ear was similar in position and height.
11	Mouth	The philtrum of the veridical was shorter and more defined, and the overall lip shape was narrower. The upper lip was thicker in the veridical, but the shape of the cupid’s bow was similar. In the veridical, the lip fissure was more irregular and the lips were more protrude.
12	Chin/Jawline	The chin was similar in length and prominence, but narrower and rounder in the veridical. Jawline comparison is problematic with the difference in the head pose, but it followed the narrowing from the gonial angle.

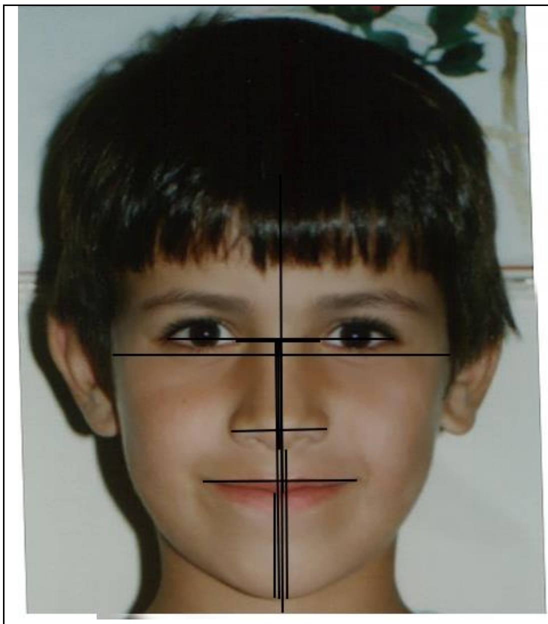
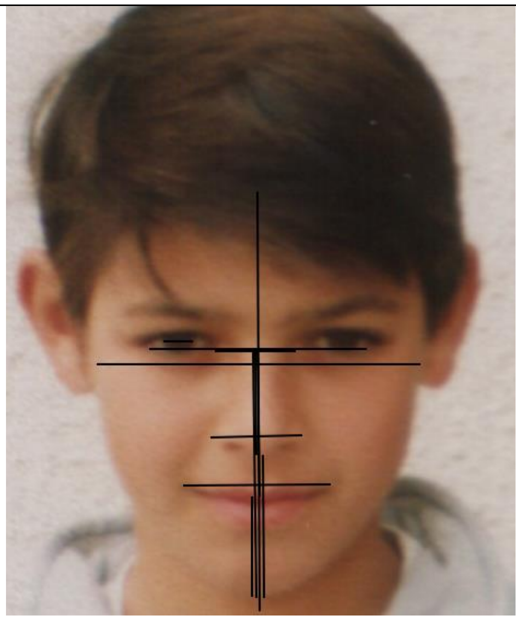
13	Neck	(not aged for FRS, therefore not comparable)
14	Facial Hair	(This was a female subject)
15	Facial Lines	The mentolabial sulcus was deeper in the veridical. The lower lid creases and the infraorbital creases were similar in position but more prominent in veridical.
16	Scars	No noticeable facial scars observed
17	Facial Marks	No obvious facial markings observed
18	Alterations	No facial alterations observed
19	Others	N/A

4.5.4.3 FM037 comparison

Table 25 compared the age progression of FM037 from age 6 years old to the target image at 13 years old. The age progression and the target image were both near frontal views. The head was tilted more downwards in the target image, and the gaze of the subject was not directed at the camera. Table 26 showed the morphological facial components comparison.

Table 25: FM037 Age 6 > 13 comparison images

		
Original age 6 years	Age progression to age 13 years	Target/veridical image at age 13 years

<table>
<tr><td></td><td></td></tr>
<tr><td>Prediction measurements</td><td>Comparing prediction to the target image</td></tr>
<tr><td colspan="2"> </td></tr>
</table>


Superimposing feature outlines of the target/veridical image onto the age progression

Table 26: FM037 age 6 > 13 Facial Components comparison

1	Skin	The veridical was darker in skin tone.
2	Face/Head Outline	The shape of the cranial vault was narrower and more oval in the veridical, especially with the lower 1/3 of the face, possibly affected by the head pose.
3	Face/Head Composition	Overall, the proportions/position of features were similar when compared to the predicted age progression measurements. However, the mouth was more inferior in the veridical; possibly affected by the head pose.
4	Hairline pattern	Hairline was masked by the difference in hairstyle.
5	Forehead	The forehead was masked by the difference in hairstyle.

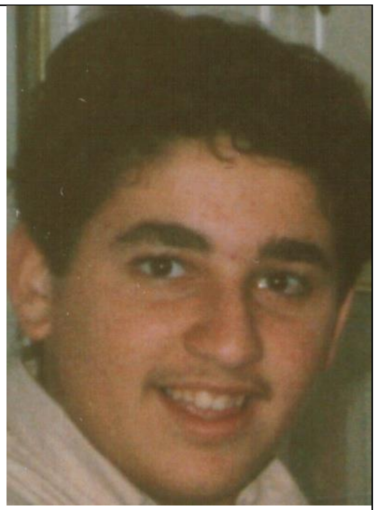
6	Eyebrows	Eyebrows were similar in position, but the left eyebrow was different in tilt and shape; possibly affected by facial expression. The medial side of the left eyebrow was closer to the left medial canthus. The hair density and distribution were similar between the age progression and the veridical.
7	Eyes	The superimposition and the predicted measurements (en-en) suggested that the iris diameter in relation to the position of the inter-eye distance were similar. Although the width was similar in relation to iris diameter, the opening eye fissure was smaller in the veridical image; possibly affected by facial expression. The shape of the inner and outer corners was similar in shape and position, but with the low image quality of the veridical, the comparison was difficult. The inferior palpebral furrows (creases) were similar in length and position; the superior palpebral furrow was not well defined in the veridical with the low image quality. The eyeball prominence was similar; more sclera was visible in the veridical. The shape and angle of the medial canthus were similar, but it was not well defined in the depiction. No obvious asymmetry observed.
8	Cheeks	The positioning of the cheekbones was similar but slimmer and flatter in the veridical.
9	Nose	In relation to other facial components, the shape, length and width of the nose were similar; the tip of the nose was more prominent in the veridical. The nasal bridge was similar in width; the nasal body was similar in width but more prominent in the veridical. The nasal tip was similar in shape but more downward pointing in the veridical. The nasal base was similar in width and height. Alae shape and thickness were similar, but the nostrils had more depth and shadows in the veridical.
10	Ears	Similar in shape and height, but with the difference in the head pose, ears were different in position.
11	Mouth	The philtrum of the veridical was longer and more defined, but the overall lip shape was similar in width and height in relation to other facial features. The irregularity and the asymmetry of the lip fissure were similar, but the lower lip was more curved and protruded in the veridical.
12	Chin/Jawline	Chin shape in both images was rounded with similar relative length and prominence, however, the jaw was narrower with less facial fat in the veridical thus gave a slimmer lower 1/3 of the face.
13	Neck	(not aged for FRS, therefore not comparable)
14	Facial Hair	(Facial hair not present)
15	Facial Lines	The mentolabial sulcus was more prominent in the veridical. The nasolabial creases were more defined and longer in the veridical.

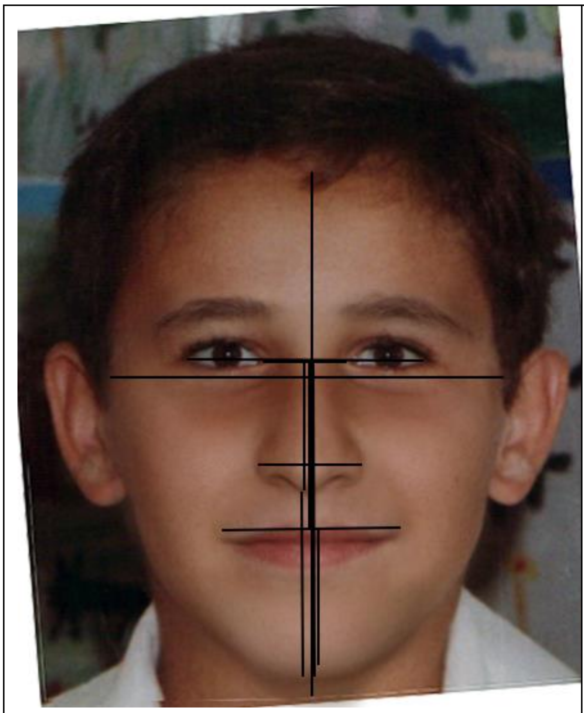
		The lower lid creases and the infraorbital creases were darker and more prominent in the veridical.
16	Scars	No noticeable facial scars observed
17	Facial Marks	No obvious facial markings observed
18	Alterations	No facial alterations observed
19	Others	N/A

4.5.4.4 FM044 comparison

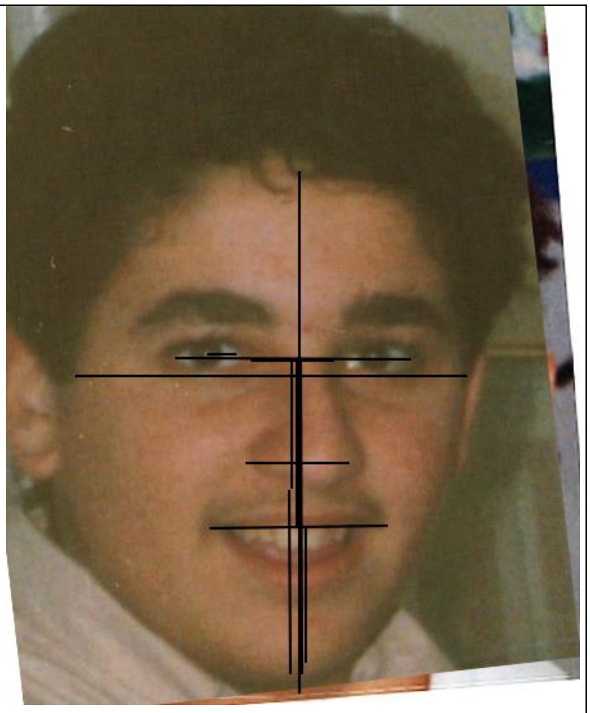
Table 27 compared the age progression of FM044 from age 5 years old to the target image at 17 years old. The age progression and the target image were both near frontal views; the head pose of the target image was angled towards the left where the right side of the face was shown more. Table 28 showed the morphological facial components comparison.

Table 27: FM044 Age 5 > 17 comparison images

		
Original age 5 years	Age progressed to age 17 years	Target/veridical image at age 17 years



Prediction measurements



Comparing prediction to the target image

Superimposing feature outlines of the target/veridical image onto the age progression

Table 28: FM044 age 5 > 17 Facial Components comparison

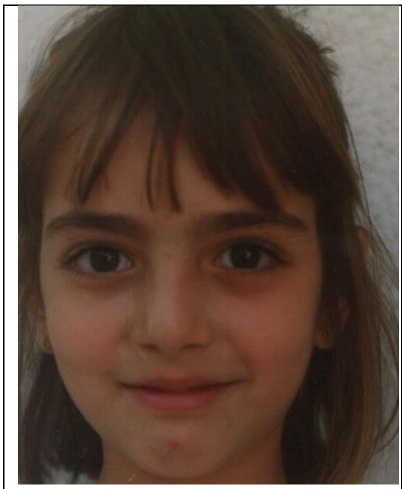
1	Skin	Overall texture and tone were similar.
2	Face/Head Outline	The shape of the cranial vault was wider and larger in the veridical, possibly affected by the head pose.
3	Face/Head Composition	The proportions/position of features were similar when compared to the predicted age progression measurements.
4	Hairline pattern	Hairline was lower in the veridical.
5	Forehead	The forehead was shorter and narrower in the veridical.
6	Eyebrows	Eyebrows were similar in position and shape, but the veridical was thicker with a higher hair density.
7	Eyes	The superimposition and the predicted measurements (en-en) suggested that the iris diameter in relation to the position of the inter-eye distance were similar. The overall shape, position of the superior palpebral furrows, eye prominence, sclera visibility, shape and angle of the medial canthus were all similar. The inferior palpebral furrows were positioned slightly lower in the veridical.
8	Cheeks	Positioning and the shape of the cheekbones were similar.
9	Nose	In relation to other facial components, the shape, length and width of the nose were similar; the tip of the nose was wider and more prominent in the veridical. The nasal bridge and the nasal body were similar. The nasal tip was similar in shape but more downward pointing in the veridical. The nasal base was similar in width and height. Alae shape was rounder and thicker in the veridical. Nostrils were not visible in the veridical.
10	Ears	Similar in position, size and shape.
11	Mouth	Taken into account the difference in facial expression (mouth opened), the philtrum was similar in width and length. The upper lip shape, cupid’s bow and thickness were similar. The lower lip was similar in thickness but more curved and protruded in the veridical.
12	Chin/Jawline	Chin shape in both images was rounded and similar in shape, but slightly more prominent and shorter in the veridical. The jawline was similar in shape but less defined in the veridical with more facial fat.
13	Neck	(not aged for FRS, therefore not comparable)
14	Facial Hair	Age progression was not presented with facial hair. Facial hair above the upper lip of the veridical was present.
15	Facial Lines	The mentolabial sulcus was more prominent in the veridical. The nasiolabial creases were straighter and deeper in the veridical (exaggerated by facial expression). The infraorbital creases were lower in the veridical. The veridical had stronger marionette lines.

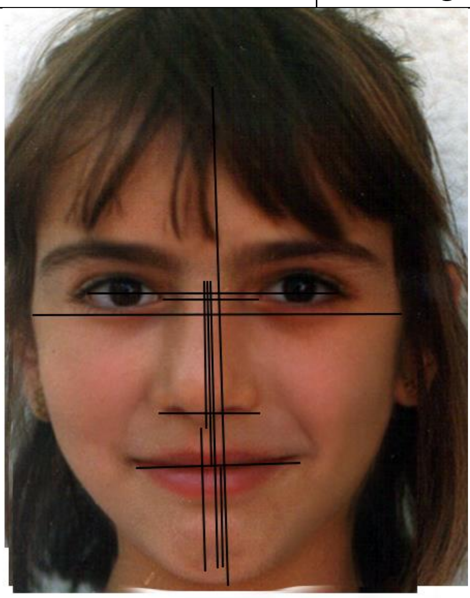
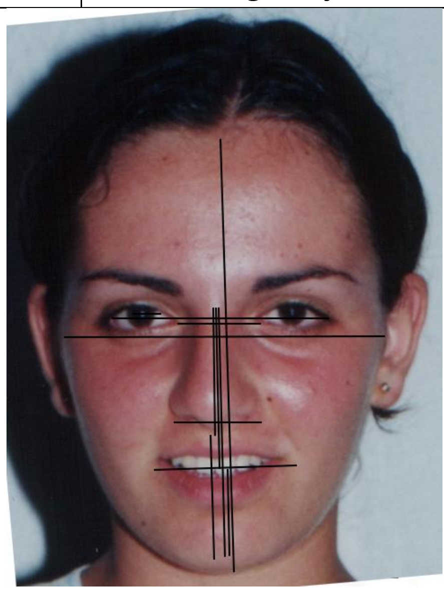
16	Scars	No noticeable facial scars observed
17	Facial Marks	No obvious facial markings observed
18	Alterations	No facial alterations observed

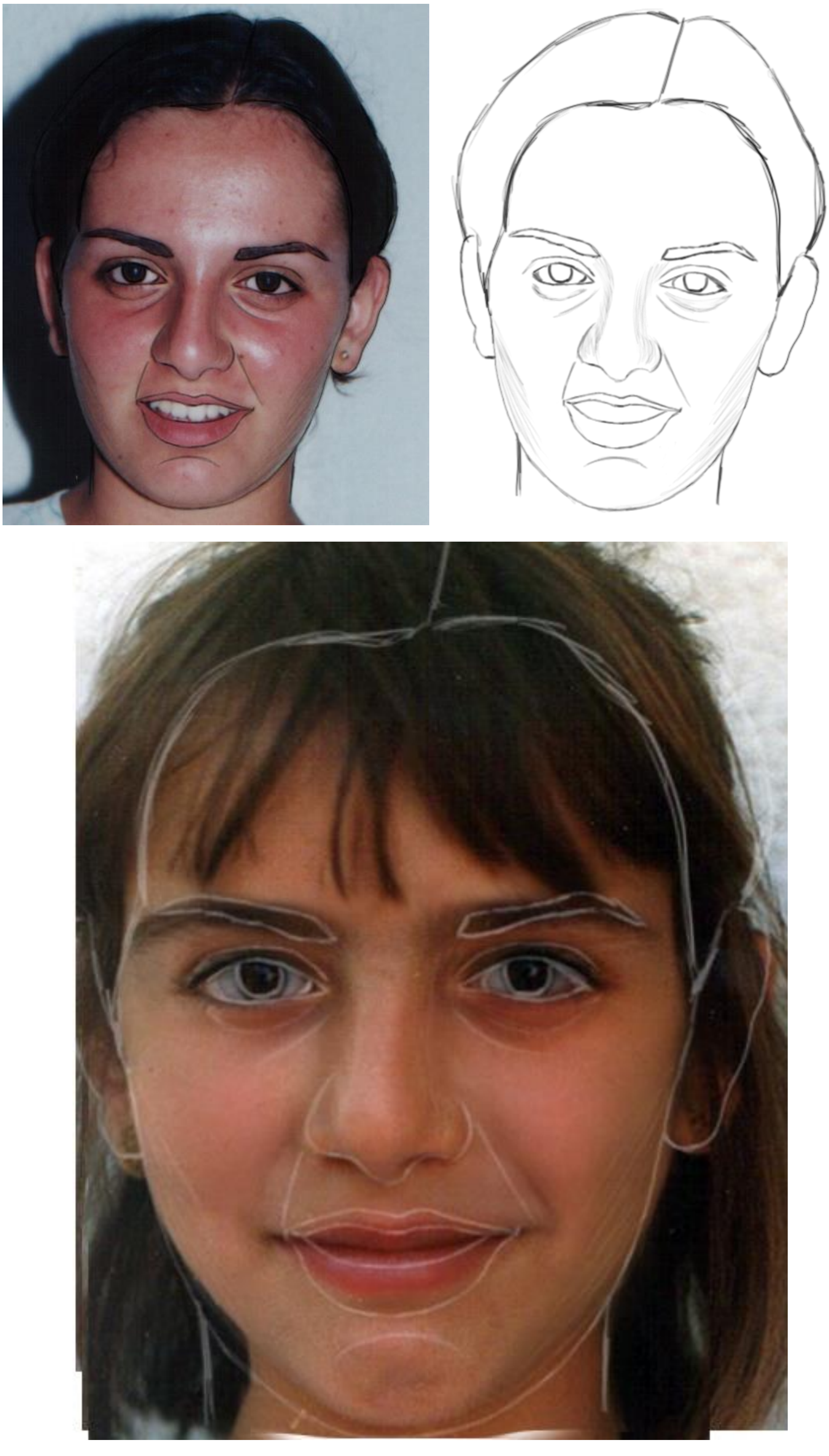
4.5.4.5 FM015 comparison

Table 29 compared the age progression of FM044 from age 5 years old to the target image at 17 years old. The age progression and the target image were both near frontal views; the head pose of the target image was angled towards the left where the right side of the face was shown more.
Table 30 showed the morphological facial components comparison.

Table 29: FM015 Age 5 > 15 comparison images

		
Original age 5 years	Age progression to age 15 years	Target/veridical image at age 15 years

	
Prediction measurements	Comparing prediction to the target image



Superimposing feature outlines of the target/veridical image onto the age progression

Table 30: FF015 Facial Components comparison

1	Skin	Overall texture and tone were different. The veridical had a warmer tone with more redness in the skin.
2	Face/Head Outline	The top of the head in the age progression was not visible, but the shape of the cranial vault was wider in the veridical.
3	Face/Head Composition	The proportions/position of features were similar when compared with the superimposition. However, the proportional measurements suggested that the nose and mouth were positioned higher in the veridical.
4	Hairline pattern	Hairline was masked by the difference in hairstyle.
5	Forehead	The forehead was masked by the difference in hairstyle, but relative to the height of the head, the width was similar in the superimposition.
6	Eyebrows	Eyebrows were similar in position but more tilted and angular in the veridical. The hair was denser but the brows were shorter and thinner in the veridical. This could be affected by alteration of the brows by hair removal.
7	Eyes	The superimposition and the predicted measurements (en-en) suggested that the iris diameter in relation to the position of the inter-eye distance were similar. However, for the best fit of superimposition in relation to other facial components, the iris diameter was slighter shorter in the veridical, thus the iris appeared to be larger in the age progression. The overall shape, eye prominence, sclera visibility, shape and angle of the medial canthus were all similar. The superior palpebral furrows were shorter and more angular in the veridical. The right inferior palpebral furrow was lower and more curved in the veridical.
8	Cheeks	Cheekbone prominence and position were similar.
9	Nose	In relation to other facial components, the length and width of the nose were similar; the width and prominence of the tip of the nose were similar. The nasal bridge was narrower and rounded in the veridical. The nasal tip was similar in shape and prominence. The nasal base was narrower in the veridical with the curve of the alae. Alae were larger and more oval in the veridical. Nostrils were narrower in the veridical following the curve of the alae.
10	Ears	Ears were hidden by hair, but the veridical had a longer left ear.
11	Mouth	Taken into account the difference in facial expression (mouth opened), the philtrum was similar in width and length. The lips were narrower in the veridical; possibly affected by facial expression. In the veridical, the upper lip was thinner and the cupid’s bow was wider. The lower lip was similar in thickness but more curved in the veridical.

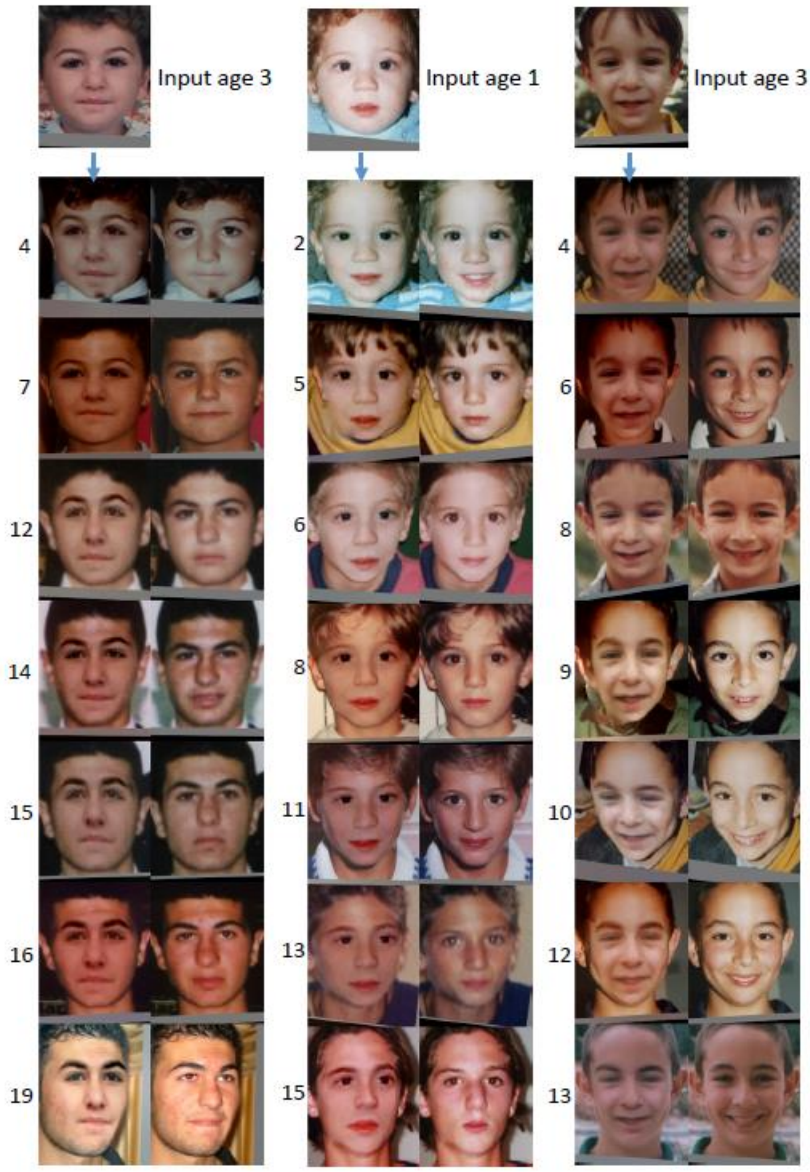
12	Chin/Jawline	Chin shape in both images was rounded with similar prominence and relative length. However, the jawline was more rounded in the veridical thus giving a rounder and less defined gonial angle.
13	Neck	(not aged for FRS, therefore not comparable)
14	Facial Hair	(This was a female subject)
15	Facial Lines	The mentolabial sulcus was similar. The nasolabial creases were closer together and straighter in the veridical (possibly affected by facial expression). The infraorbital creases were more prominent in the veridical.
16	Scars	No noticeable facial scars observed
17	Facial Marks	Freckles were observed on the forehead of the veridical, but it was concealed by a hair in the age progression. Moles approx. 0.5mm left of the alar, and the lower right side of the face approx. 3 cm diagonally below the lips were observed. Both moles were not observed from the original, thus not depicted in the age progression.
18	Alterations	Eyebrows were most likely treated and reshaped by hair removal.
19	Others	N/A

4.5.5 Common differences

Overall, the most consistent differences were that the veridical was often presented with more tilted eyebrows and a more prominent nose. With a difference in head-pose, some comparisons were more challenging than others. Human recognition differs to machine-based recognition, although the similarities and differences of the most ‘dissimilar’ faces were addressed with the manual facial comparison, it is unclear how this differs to the recognition made by Microsoft Face API. The overall similarities using manual facial comparison was not considered, as the researcher already knew the identity. This type of comparison is a good method for practitioners to evaluate and improve their own method of manual age progression.

4.6 Experiment 2C: Manual age progression versus machine-based study

Based on the FG-NET images used in the machine-based method reported in the supplementary material of Kemelmacher-Shlizerman et al. (2014) (Figure 48), this study compared their images with the guided manual method described above. Within the supplementary material, the authors showed images of machine-based progression using subjects from the FG-NET, and these images were progressed to several different ages according to the availability of the 'ground-truth' images (i.e. images of the same individual at other different ages) within the FG-NET database.



E

Figure 48: Supplementary material in Kemelmancher-Shlizerman et al. (2014).

© 2014 IEEE

Five original (outdated/input) images used for the machine-based progressions were matched to the FG-NET database. Based on the 'ground-truth' image used, these images were manually progressed to subsequent ages shown in the Kemelmacher-Shlizerman et al. (2014) supplementary material. These 'ground-truth' images will be referred to **Target images**.

The original (out-dated) images, the manual age progressions and the machine-based progressions will be compared to the target image from the FG-NET database using the Microsoft Face API V1.0. This produces a score of likeness between two images for an objective comparison.

Two different types of machine-based age-progressed image were available for comparison (Figure 49), the paper referred to these images as 'relightable average images' where the age-progressed image could be re-illuminated from any direction with realistic shadow effects to match the input image. The left image in Figure 49 is the 'unblended original cropped' image, these images were referred to as **Cr**. The image on the right has the same illumination as the target image (i.e. blended into the ground-truth head). These 'illumination aware' images were referred to as **IA**.

Figure 49: Age progression image sets from Kemelmacher-Shlizerman et al. (2014)

One of the concerns in using the images derived from Kemelmacher-Shlizerman et al. (2014) was that the 'illumination aware method' was based on the 'ground-truth' images with similar texture. Although the internal features of the face predicted by the algorithm differed to the 'ground-truth', skin texture and lighting were matched; the external information of the image (i.e. hair and clothing) remained identical. Displaying results in the style as shown in Figure 48 could bias a likeness between the two images. Human recognition here would be unfamiliar face identification, which could be affected by the external features of the subject. The same image content will also bias the algorithm to produce a higher similarity rating.

One disadvantage in using a commercial software (Microsoft Face API) would be not knowing exactly how the algorithm works, what parameters the algorithm is using to identify two images as the same person.

4.6.1 Method of analysis

Based on the machine-based age progression images from the supplementary material of Kemelmacher-Shlizerman et al. (2014), five male subjects were identified as M031, M069, M074, M079 and M080 from the FG-NET database. The 'Illumination aware' (**IA**) technique was performed on all five subjects, but the 'unblended original cropped' (**Cr**) images were only shown on three subjects (M069, M079 and M080) within the supplementary material.

Each subject had one original image, and that same image was manually age-progressed 6-7 times using the Method of manual age progression, (p.72). The age separation of the progressions was based on the published images from the **KS** study (i.e. Kemelmacher-Shlizerman et al. (2014)). Progressions were limited to faces between ages 1-18 years old due to the constraints of the manual method.

The original images were compared to the target images. Microsoft API produced a confidence score between 0-1, and scores were documented in Excel. When the confidence score was above 0.5, Microsoft Face API suggested, "*The two faces belong to same person*"

The manual age progressions were compared to the original image and the target image using Microsoft API.

Both 'IA' and 'Cr' were extracted from the KS study and compared to the original image and the target image using Microsoft API. The external-features/background of KS(IA) was the same as the target image, 'IA-Cr' from subject M069, M079 and M080 were cropped to the same size as the Cr for comparison The manual age progressions were compared to the KS(IA) and KS(Cr) using Microsoft API. Manual Age progressions were cropped to the same size as Cr for comparison.

All comparisons with the target images (i.e. Original, Manual and Machine-based age progressions) indicated the best-performed method. When compared to the target image, the correlation between the age progressions (manual and machine-based) with the original photograph was explored. The likeness between the age progressions and the original photograph had a correlation to the ‘success’ rate (i.e. comparison to target image) was explored, and data were examined with univariate analysis of variance using SPSS.

4.7 Experiment 2C: Results

Five male subjects (M031, M069, M074, M079, and M080) from the FG-NET were manually progressed to various different ages (Figure 50). The age of the original image differed between subjects. The simulated age separation was also slightly different between the subjects. The age progressions were chosen to match the age simulated from Kemelmacher-Shlizerman et al. (2014) as references. A list of abbreviations is shown in Table 31.

Table 31: relevant abbreviation(s)

Abbreviation(s)	
AP	Age progression
Manual AP(Cr)	Manual AP cropped to the same size as KS(Cr)
T-Manual AP	Manual AP transformed to the same scale, rotation and position as the cropped target image
KS(IA)	Kemelmacher-Shlizerman et al. (2014) Illumination Aware blended into the ground-truth head
KS(Cr)	Kemelmacher-Shlizerman et al. (2014) Unblended original cropped
KS(IA-Cr)	KS(IA) cropped to the same size as KS(Cr)
T-KS(IA)	KS(IA) transformed to the same scale, rotation and position as the cropped target image
T-KS(IA-Cr)	KS(IA-Cr) transformed to the same scale, rotation and position as the cropped target image

KS(IA) and Manual AP from M069, M079 and M080 were cropped to the same size as KS(Cr)

Overall, 34 manual age progressions were produced (Figure 50). There were no overlaps of depiction between Experiment 2C and 2B. The original age of the 5 subjects ranged from 1-3 years old, and the age gap of the progression within each subject ranged from 1-14 years. When the confidence score is above 0.5, Microsoft Face API will suggest, "*The two faces belong to same person*"

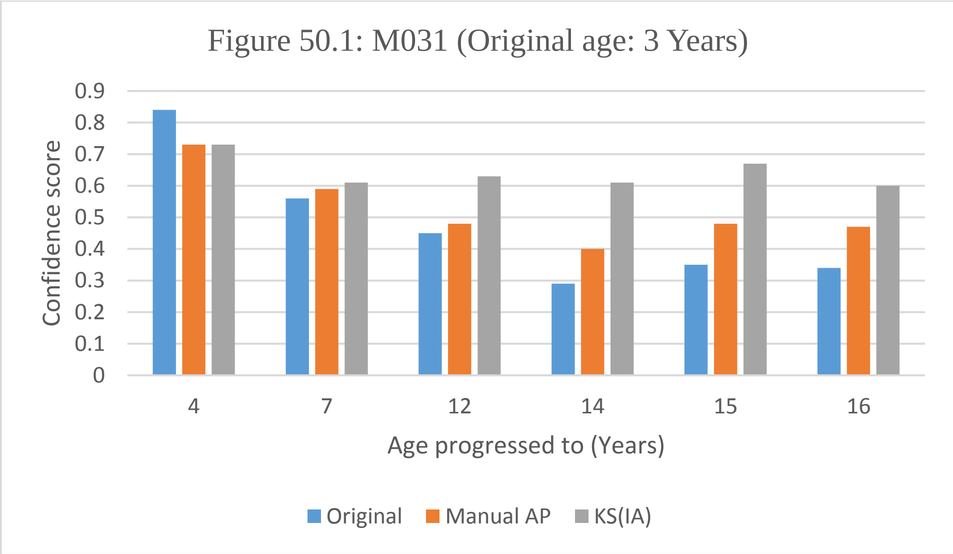


Figure 50.1: M031 (Original age: 3 Years)
Confidence score
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
4
7
12
14
15
16
Age progressed to (Years)
Original
Manual AP
KS(IA)

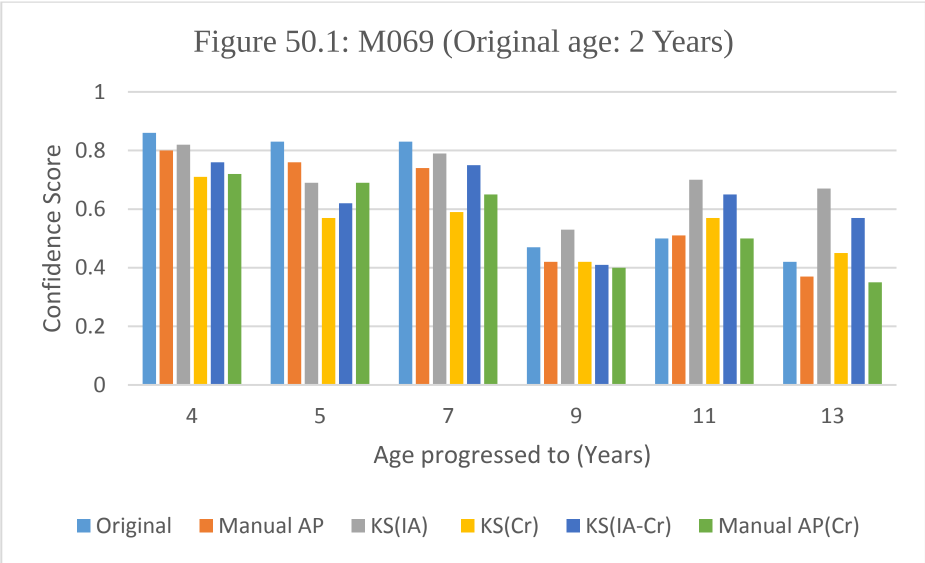


Figure 50.1: M069 (Original age: 2 Years)
Confidence Score
1
0.8
0.6
0.4
0.2
0
4
5
7
9
11
13
Age progressed to (Years)
Original
Manual AP
KS(IA)
KS(Cr)
KS(IA-Cr)
Manual AP(Cr)

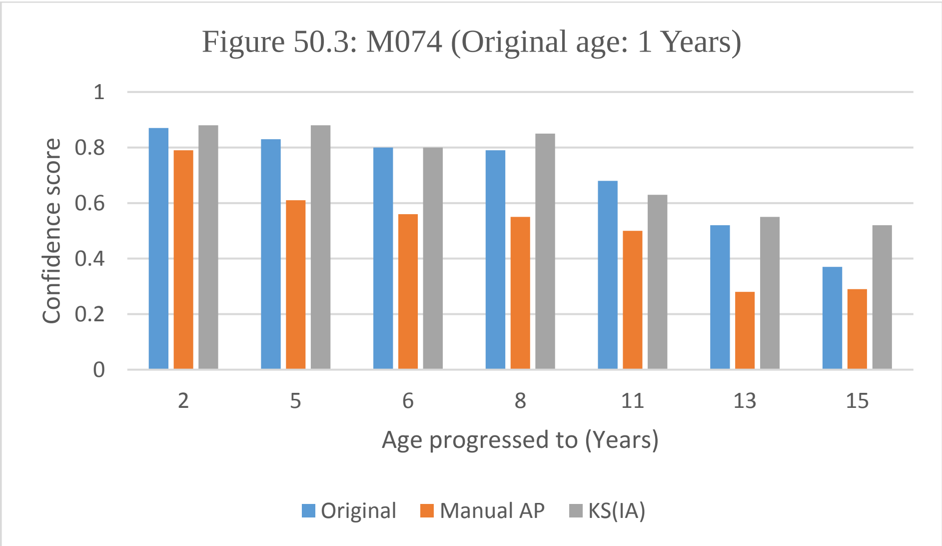
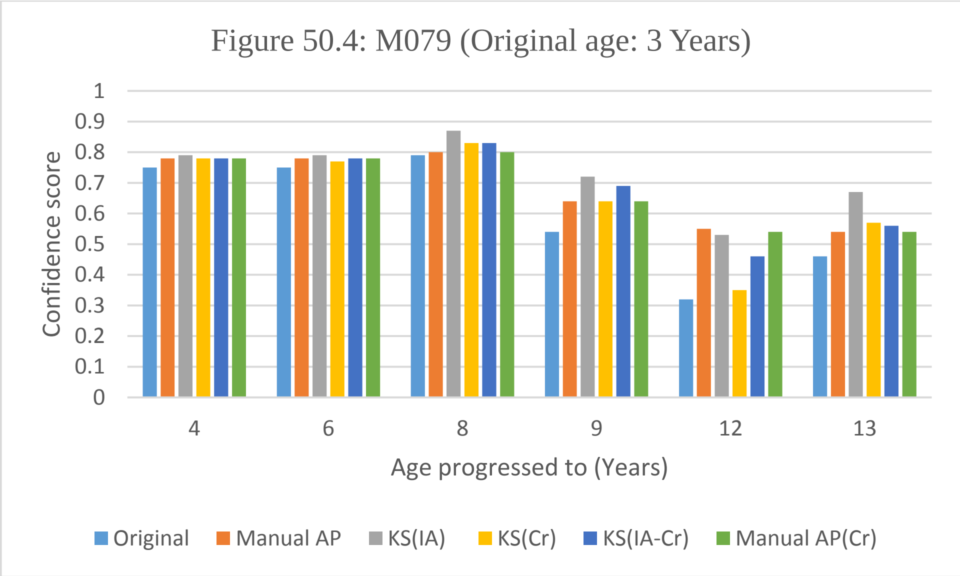


Figure 50.3: M074 (Original age: 1 Years)
Confidence score
1
0.8
0.6
0.4
0.2
0
2
5
6
8
11
13
15
Age progressed to (Years)
Original
Manual AP
KS(IA)



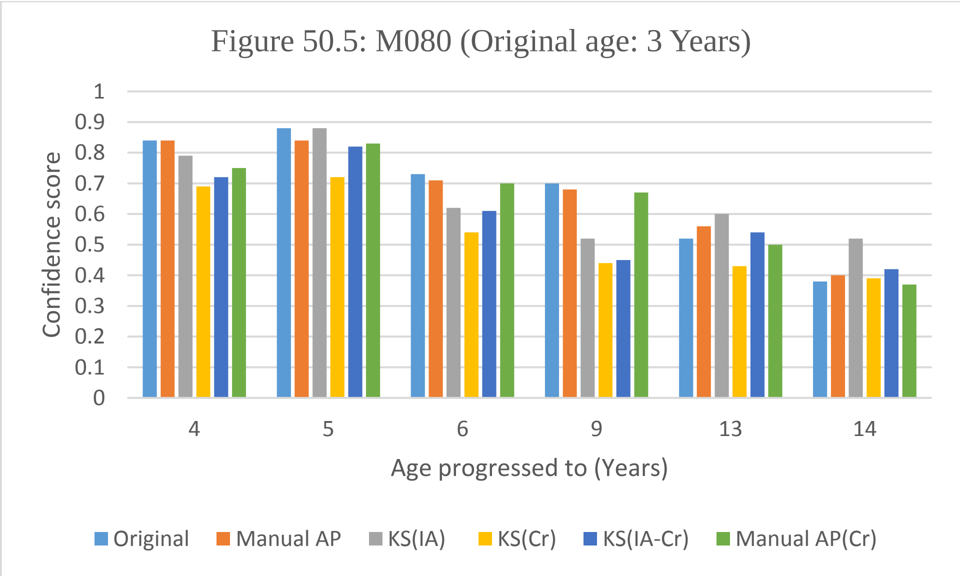


Figure 50: Recognition rates when compared to the target image for different age progression methods and the original image

4.7.1 Addressing bias

This section explored how the Microsoft API can be affected by different image conditions

- Section 4.7.1.1 explores the bias generated by the similarity of external background
- Section 4.7.1.2 explores the effect of image transformation
- Section 4.7.1.3 explores the effect of cropped images (i.e. facial completeness)

4.7.1.1 Comparing the machine-based conditions

To establish if the external information of the illumination aware images (IA) had a bias effect on the algorithm, KS(IA) was cropped to the same size as the original cropped images (Cr). The cropped KS(IA) are displayed as KS(IA-Cr).

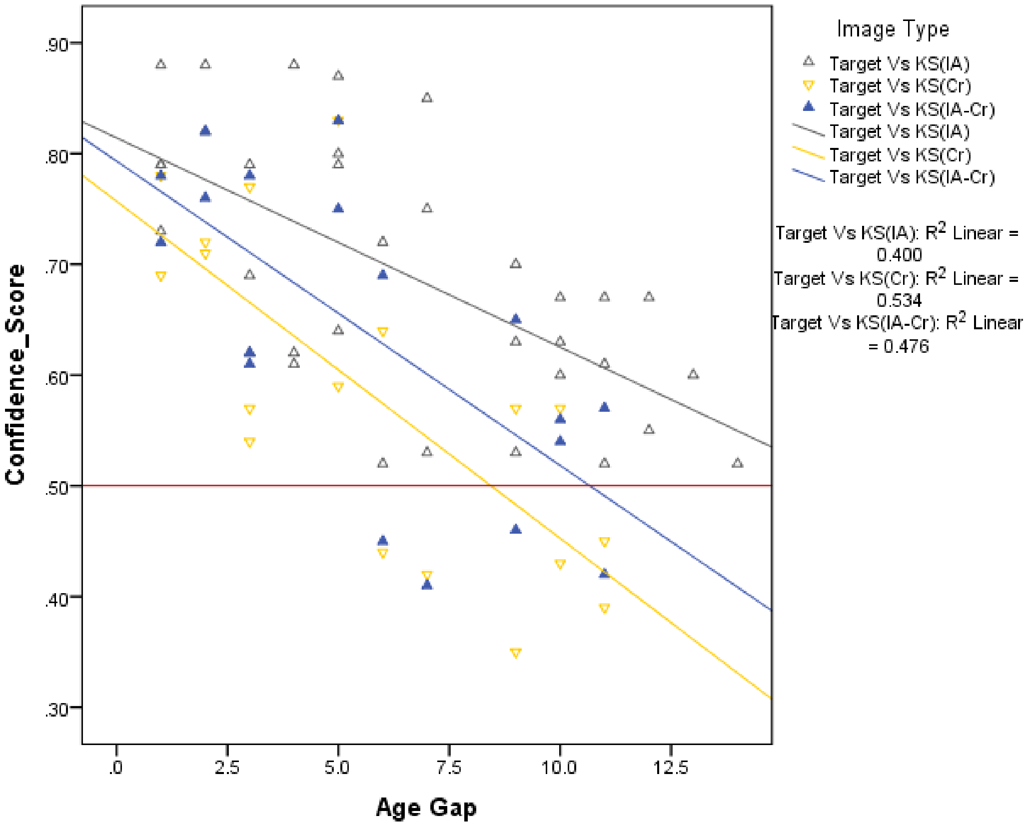


Figure 51: Face recognition rate comparing the different KS cropped conditions in relation to age gap (years)

Figure 51 showed the confidence scores for the machine-based method of the unblended original cropped (Cr), the illumination aware uncropped images (IA) and the illumination aware cropped images (IA-Cr) (see Appendix 3 for the statistics report).

There was a significant difference between the three progression types [$F(2, 33.491) = 33.121$, $p < 0.001$]. Post hoc comparison using the Tukey HSD test indicated that the illumination aware uncropped images (IA) ($M=0.694$, $SD=0.121$) was significantly different from the unblended original cropped (Cr) ($M=0.581$, $SD=0.147$). However, the illumination aware cropped images (IA-Cr) ($M=0.634$, $SD=0.142$) did not significantly differ from the other conditions.

This suggests that the external-information/background or the 'completeness' of a face does have an effect on this algorithm for face recognition. However, this difference was not significant.

The illumination aware cropped images (IA-Cr) showed a higher confidence score when compared to the unblended original cropped (Cr) and this suggests having the 'correct' texture/illumination does have an effect on this algorithm for face recognition. However, this difference was not significant.

4.7.1.2 The effect of image transformation

To establish if the 'completeness' of a face had an effect on this algorithm for face recognition, the effect of image transformation (translation, rotation and scale) was first considered. Images used in this test were cropped to the same size with the face at the same position. For example, KS(IA) was cropped to the same size as the target images where the face was in the same position. T-KS(IA-Cr) was the transformed set of KS(IA-Cr).

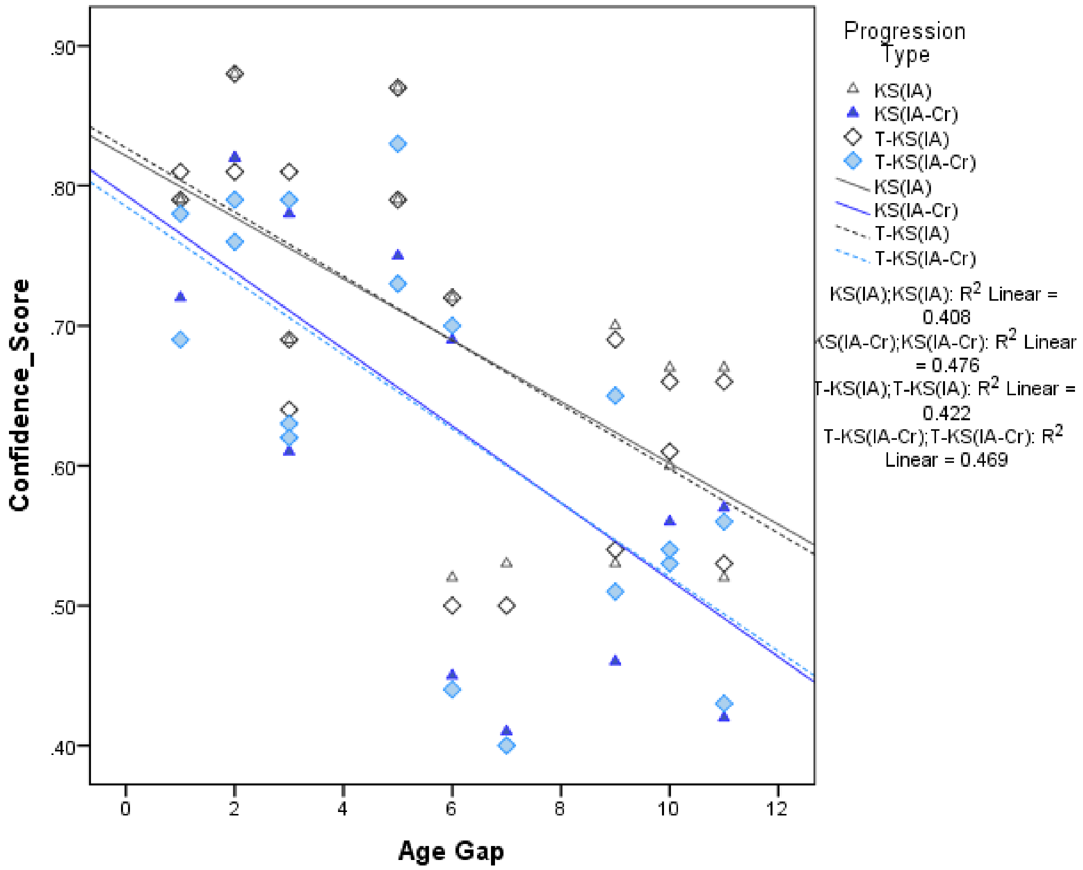


Figure 52: Face recognition rate comparing the transformed set of KS cropped conditions in relation to age gap (years)

The confidence score of the transformed images was shown alongside the original datasets as dotted lines in Figure 52 (see Appendix 3 for the statistics report).

- The effect of transformation between T-KS(IA) (M=0.694, SD=0.124) and KS(IA) (M=0.694, SD=0.121) was not significant [t(34)=0.000, p = 1.000].
- The effect of transformation between T-KS(IA-Cr) (M=0.632, SD=0.136) and KS(IA-Cr) (M=0.634,SD, 0.140) was not significant [t(34)=0.048, p = 0.962].
- Although the image of the face remained the same, with a change in position of the image, confidence scores were not identical and showed a slight difference between the datasets. However, these differences were not significant.

4.7.1.3 Facial completeness

Does the 'completeness' of a face have an effect on this algorithm for face recognition?

When the cropped and the full image of the age progressions types were compared to the target image, the machine-based method (KS(IA)) showed a higher score difference when compared to the manual method (MAP) (Figure 53). This difference was significant between KS(IA) (M=0.06, SD=0.03343) and MAP (M=0.0283, SD=0.03330) conditions; [t(34)=2.847, p=0.007]. The background information of the manual method was different to the target, and the background information was the same in the machine-based comparisons. When the background information was taken away, the FRS suggests the machine-based method was more different. The confidence score between the KS(IA) and KS(IA-Cr) had a maximum difference of 0.12, this figure was higher than the Manual AP set. Although the inner face area between the two sets were identical, there was a difference between the manual and automated conditions. This suggests the background information could have contributed to a higher recognition rate.

In addition, the confidence score between the cropped and the original versions of the age progression was compared, and the score was derived by deducting the confidence score from 1.0, represented as 1-MAP and 1-KS(IA) in Figure 53. Since the images were the same with just a comparison without the background information, if the algorithm had only considered the internal face, the confidence score should have remained identical with a score of 1.0. The results suggests otherwise.

When the original age progression was compared to the cropped stimuli (1-KS(IA) and 1-MAP), none achieved a confidence score of 1.0, this suggests the algorithm was affected by the completeness of the face. There was a significant different between the KS(IA) (M=0.0567, SD=0.02249) and the MAP (M=0.0406, SD=0.01798) conditions; [t(34)=2.347, p=0.023]. The difference between the two conditions ranged from 0.01 to 0.1. The background in this comparison was not identical, this may suggest the difference may not be all related to the similarity in background, but the actual image itself, for example, the parameter of the crop.

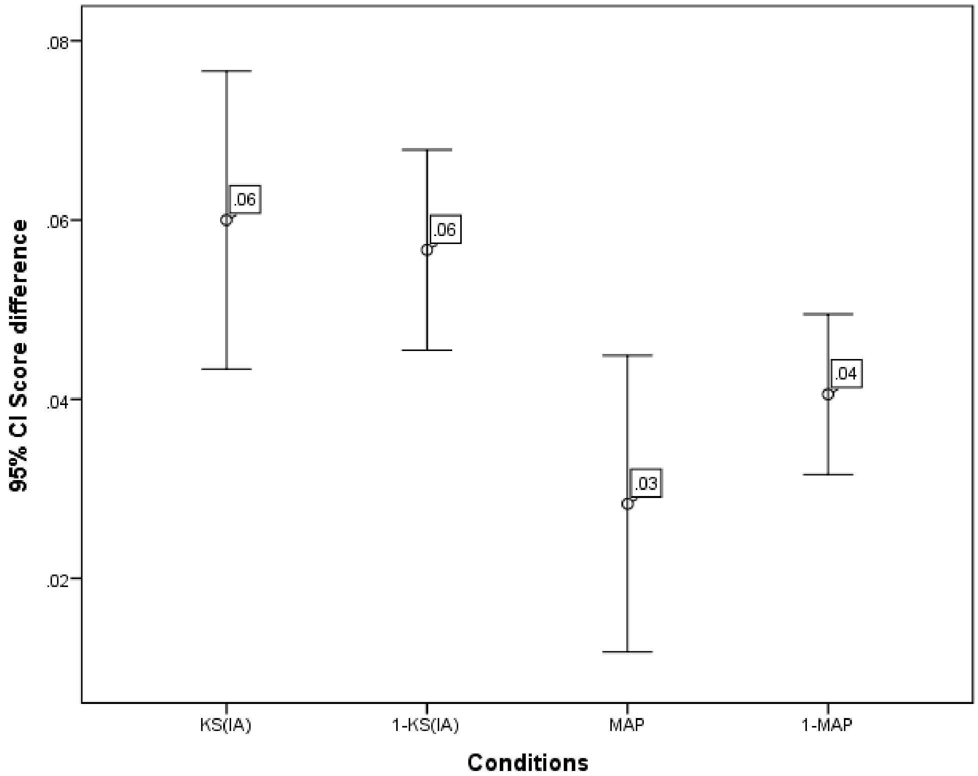


Figure 53: Comparing facial completeness between the machine-based and manual age progressions *(1- represents confidence score difference deducted from 1.0)

These results suggests the external-information/background or the ‘completeness’ of a face could have caused the difference in face recognition using the Microsoft API.

4.7.2 Original & Manual age progression versus target

Figure 54 compared the original image and the manual age progression

- 23/34 (67.7%) original images were recognised by the algorithm as the same individual
- 24/34 (70.6%) manual age progressions were recognised by the algorithm as the same individual

The recognition rate of the original images generated a similar trend when compared to the manual age progressions (Figure 54)

  - 19/34 (55.9%) original images achieved equal (n=1) or higher (n=18) recognition rate when compared to the manual age progressions.
  - 16/34 (47.1%) manual age progressions achieved equal (n=1) or higher (n=15) recognition rate when compared to the original images

Recognition rate decreased as the age gap of the progression increased and both trends fell below recognition (0.5) from the Microsoft Face API with an age gap around 10 years (see Figure 54 and Appendix 3).

When compared to the target, the original image (M=0.621, SD=0.194) was not significantly different to the Manual AP (M=0.594, SD=0.164) [t(66)=0.609, p=0.545].

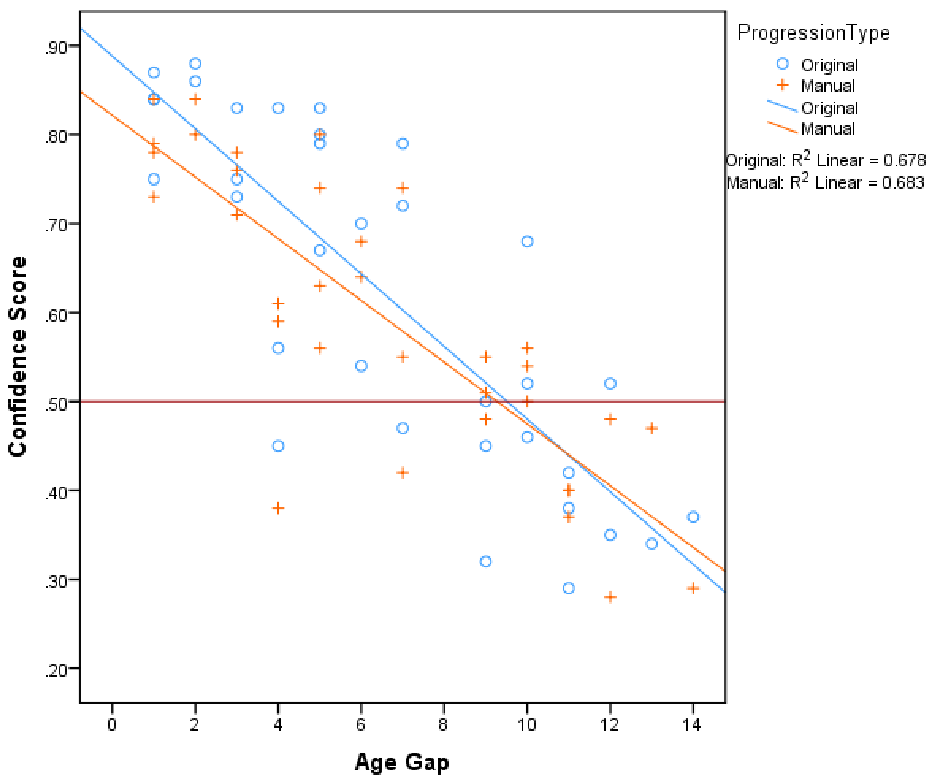


Figure 54: Face recognition rate in relation to age gap (years), comparing the progression types (original and manual age progression) versus the target image

4.7.3 Original, KS(Cr) and KS(IA-Cr) versus target

For recognition scores of the original images, see section 4.7.2. Figure 55 compares the original to the unblended original cropped images KS(Cr) and the illumination aware cropped images KS(IA-Cr).

- 12/18 (66.7%) KS(Cr) were recognised by the algorithm as the same individual, but had a lower recognition rate in comparison to the original images.
- 14/18 (77.8%) KS(IA-Cr) were recognised by the algorithm as the same individual, and had a higher recognition rate in comparison to the original images.
- Recognition rate decreased as the age gap of the age progression increased

The recognition rate of the KS(Cr) generated a similar trend when compared to the original images and the manual age progressions (Figure 55)

- o 9/18 (50%) KS(Cr) achieved equal (n=0) or higher (n=9) recognition rate when compared to the original image
- o 9/18 (50%) original image achieved equal (n=0) or higher (n=9) recognition rate when compared to the KS(Cr)
- o 8/18 (44.4%) KS(IA-Cr) achieved equal (n=0) or higher (n=8) recognition rate when compared to the original image
- o 10/18 (55.6%) original image achieved equal (n=0) or higher (n=10) recognition rate when compared to the KS(IA-Cr)

The Ks(Cr) trend fell below the recognition from the Microsoft Face API with an age gap around 8-9 years, and 11 years for KS(IA-Cr)

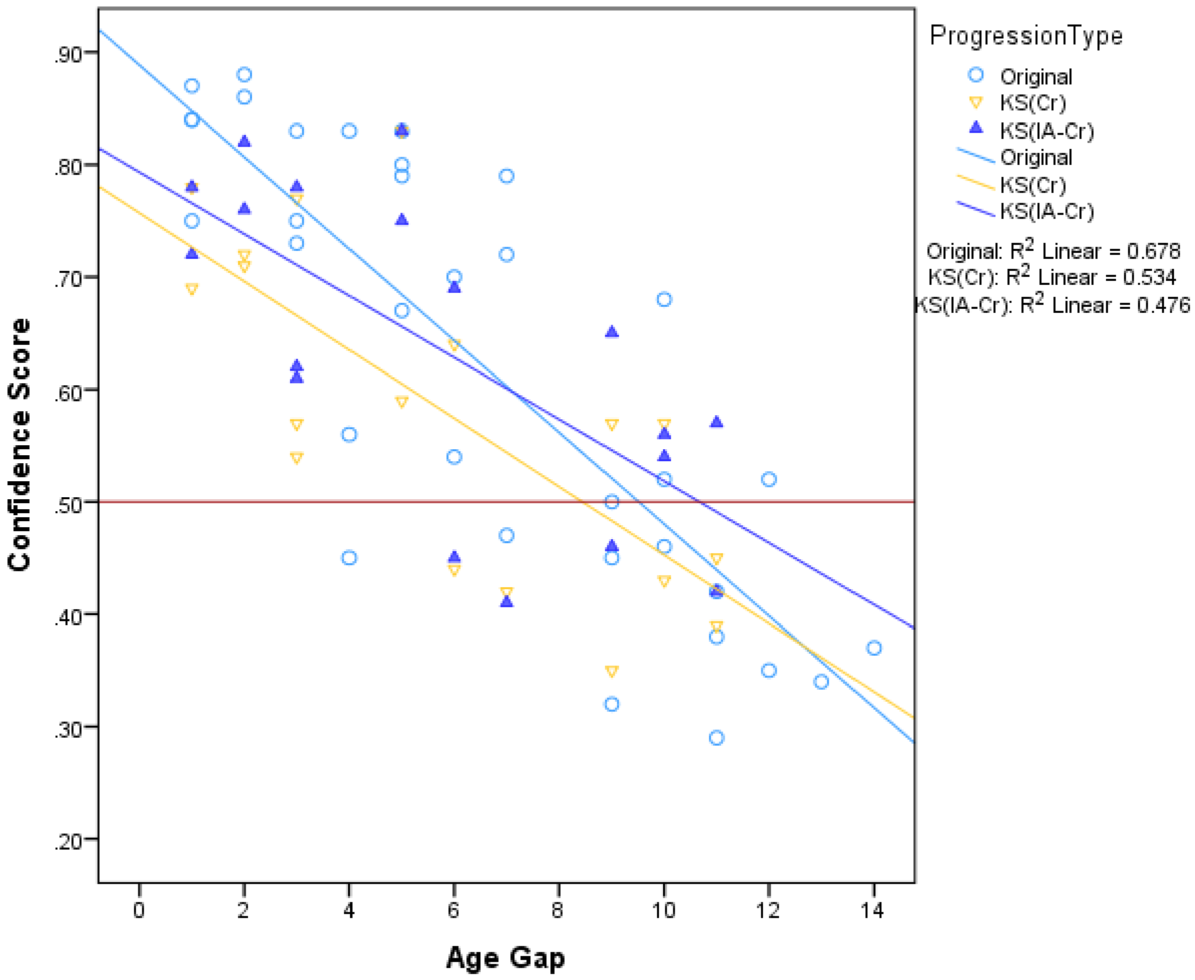


Figure 55: Face recognition rate in relation to age gap (years), comparing the progression types (Original, KS(Cr) and KS(IA-Cr)) versus the target image

When compared to the target, the original image (M=0.621, SD=0.194) was not significantly different to KS(Cr) (M=0.581, SD=0.147); [t(43.641)=0.823, p=0.415], or KS(IA-Cr) (M=0.634, SD=0.140) [t(44.967)=-0.296, p=0.769]. Appendix 3 details the statistics report.

Interestingly, there was a slight significant difference between the three conditions [F(2, 17.948) = 3.747, p =0.044]. However, a post hoc comparison using the Tukey HSD test indicated that the conditions did not significantly differ from the each other. This suggests although there is a difference, the original image performed at a similar rate to the machine-based conditions.

4.7.4 Original, Manual AP and KS(IA) versus target

For the recognition scores of the original image see section 4.7.2. Figure 56 compares the original, manual AP and machine-based illumination aware KS(IA) images.

- 34/34 (100%) KS(IA) were recognised by the algorithm as the same individual
- Recognition rate decreased as the age gap of the progression increased
- In comparison to others images, KS(IA) was recognised more frequently with the increasing age gap.
    - 25/34 (73.5%) KS(IA) achieved equal (n=2) or higher (n=23) recognition rate when compared to the original images
    - 11/34 (32.35%) original images achieved equal (n=3) or higher (n=8) recognition rate when compared to the KS(IA)
    - 29/34 (85.29%) KS(IA) achieved equal (n=1) or higher (n=28) recognition rate when compared to the Manual AP
    - 6/34 (32.35%) Manual AP achieved equal (n=1) or higher (n=5) recognition rate when compared to the KS(IA)

All progressions from the machine-based method KS(IA) were recognised above a confidence score of 0.5 by the Microsoft Face API (Figure 56). This suggests the progressions generated by KS(IA) were more similar to the target images. With the bias addressed in section 2C-1 above, results displayed here must be treated with caution.

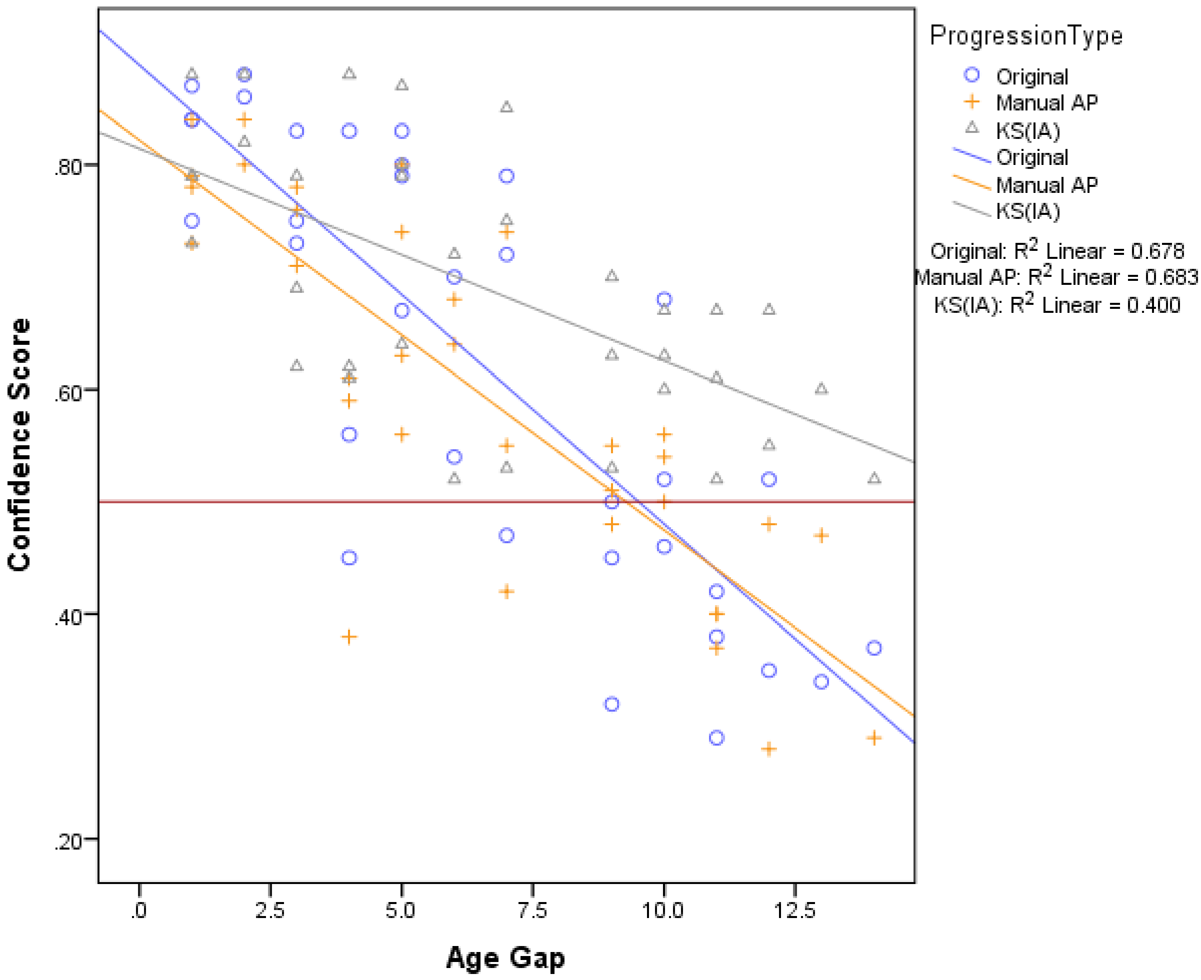


Figure 56: Face recognition rate in relation to age gap (years), comparing the progression types (Original, Manual and KS(IA)) versus target

When compared to the target, the original images (M=0.621, SD=0.194) were not significantly different to the KS(IA) (M=0.69-, SD=0.117) [t(54.170)=-1.799, p=0.078]. However, the manual AP (M=0.594, SD=0.164) was significantly different to the KS(IA) (M=0.690, SD=0.117) [t(59.608)=-2.789, p=0.007]. Appendix 3 details the statistics report.

4.7.5 Manual age progressions, KS(Cr) and KS(IA-Cr) versus target

Figure 57 indicates the recognition rate of the different age progression methods. The results are as follows:

- Recognition rate decreased as the gap between the original age and the target age increased
- 14/18 (77.8%) manual age progressions achieved equal (n=3) or higher (n=11) recognition when compared to the machine-based original cropped images KS(Cr)
- 11/18 (61.1%) manual age progressions achieved equal (n=2) or higher (n=9) recognition when compared to the machine-based illumination aware cropped images KS(IA-Cr)

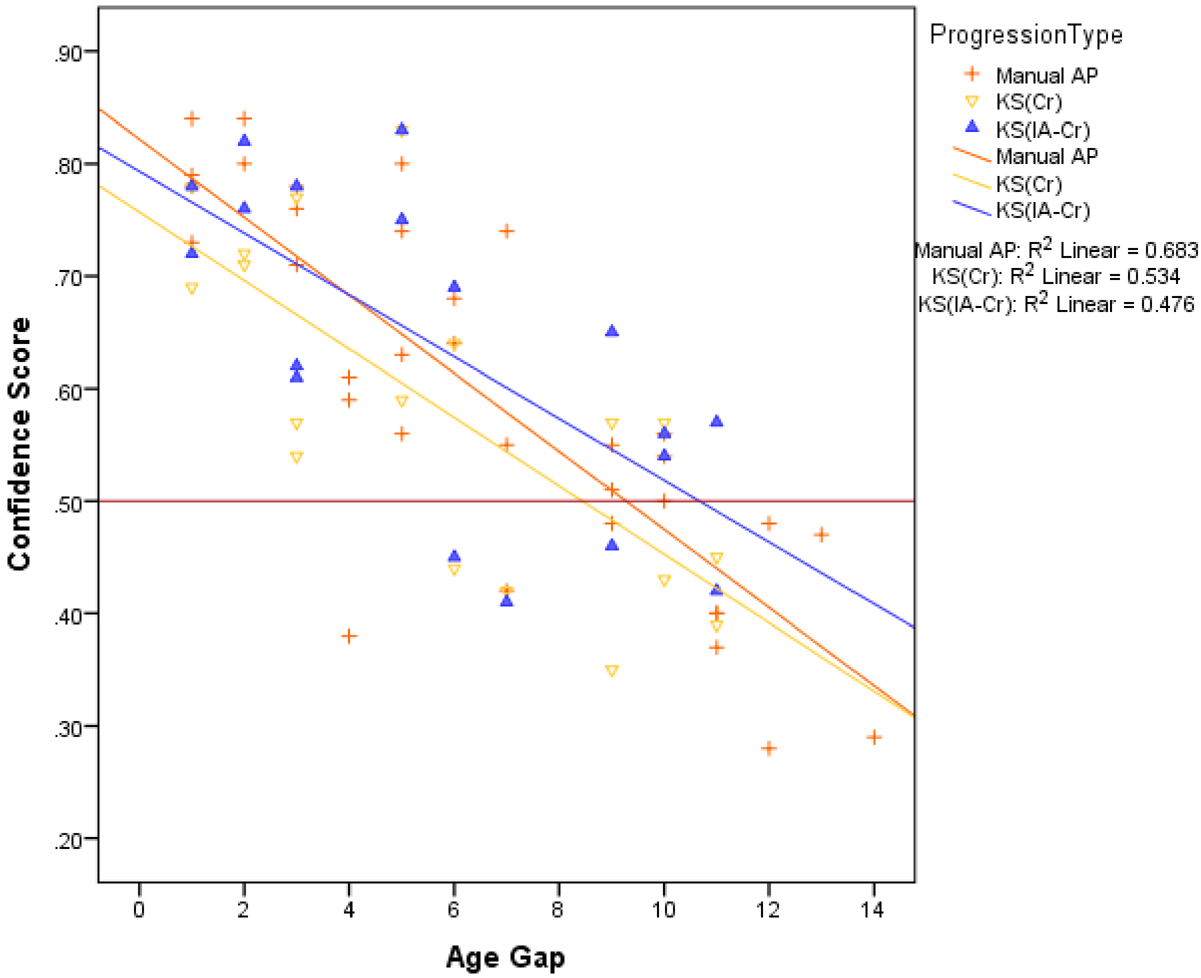


Figure 57: Face recognition rate in relation to age gap (Years), comparing the progression types (Manual, KS(Cr) and KS(IA-Cr)) versus target

When compared to the target images the manual APs (M=0.594, SD=0.164) were not significantly different to the machine-based original cropped images KS(Cr) (M=0.581, SD=0.147): t(50)=0.282, p=0.779, or the machine-based illumination aware cropped images KS(IA-Cr) (M=0.634, SD=0.140) [t(50)=-0.885, p=0.380]. There was a significant difference between the three conditions [F(2, 18.661) = 4.022, p =0.035]. However, Post hoc comparison using the Tukey HSD test indicated that the conditions did not significantly differ from the each other. This suggests although there is a difference, manual AP performed at a similar rate to the machine-based conditions (see Appendix 3 for the statistics report).

4.7.6 ALL progression types versus original

The different types of age progressions were compared to the original image in Figure 58.

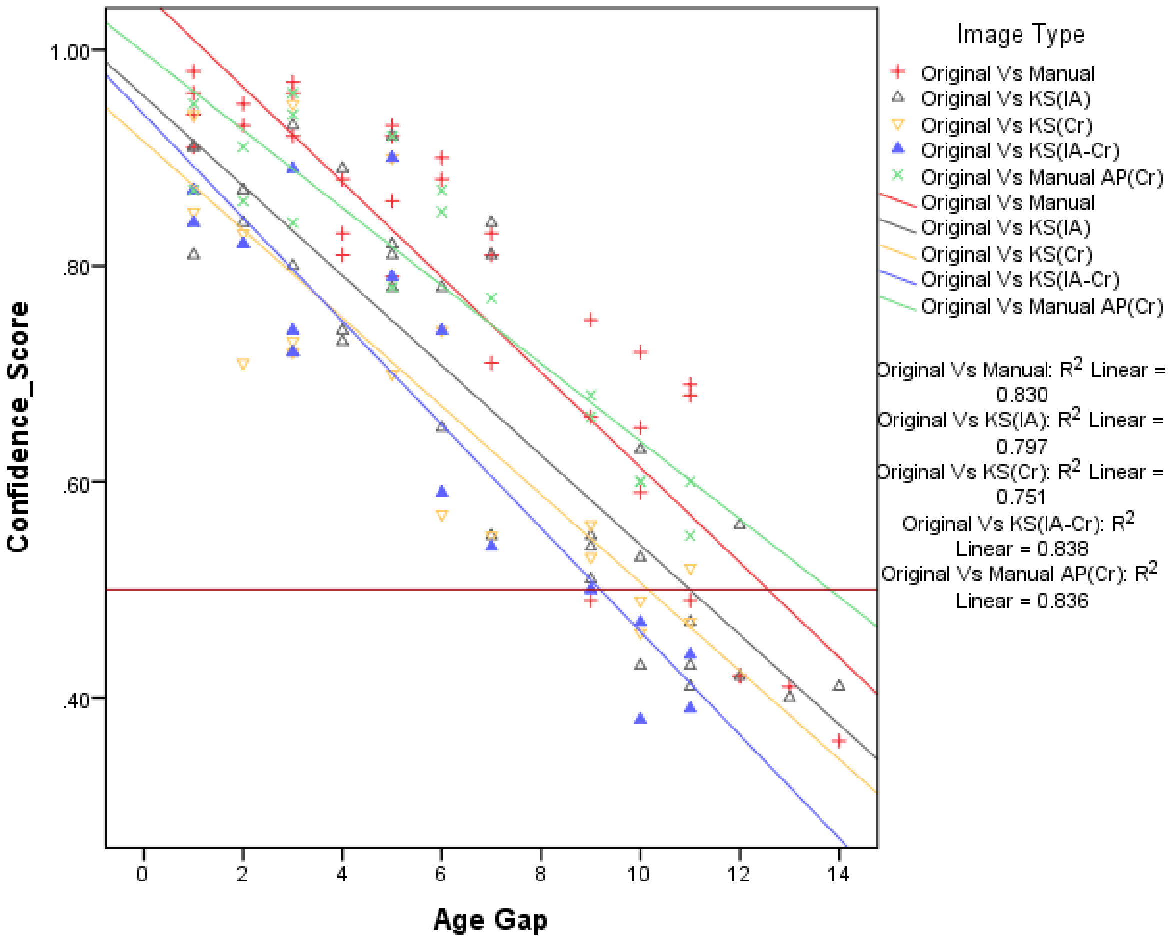


Figure 58: Face recognition rate in relation to age gap (years), comparing all progression types versus the original image

There was a significant difference between all image types [$F(4,46.169) = 17.869$, $p<0.000$] and between age gap [$F(12,44.477) = 54.846$, $p<0.000$].

Figure 58 indicated the similarities between the face recognition rates of the progression types and the original images. Post hoc comparison using the Tukey HSD test indicated that the Manual method (Manual AP) ($M=0.765$, $SD=0.188$) and its cropped version (Manual AP(Cr) ($M=0.781$, $SD=0.018$) did not record significantly different recognition rates. Similarly, the machine based method (KS(IA)) ($M=0.6844$, $SD=0.1816$) and its cropped versions KS(Cr) ($M=0.679$, $SD=0.166$), and KS(IA-Cr) ($M=0.663$, $SD=0.184$) showed no significant difference in recognition rates (see Appendix 3 for the statistic report).

Both manual conditions recorded significantly different recognition rates to the three machine-based KS conditions, and this suggests that manual age progression was more similar to the original. As the age gap increased, the likeness between the manual age progression and the original image decreased. This suggests as the age gap increases, the age progressions will be more dissimilar to both the original and the target image.
When compared to the Manual AP, the machine-based conditions decreases in confidence score with the increasing age gap. This suggests the manual age progressions became more dissimilar to the machine-based age progressions with the increasing age gap.

4.7.7 Progression types: Original or Target

The clear decrease in confidence score with the increasing age gap suggested a comparison of the age progression methods to the target and the original image (see Figure 59).

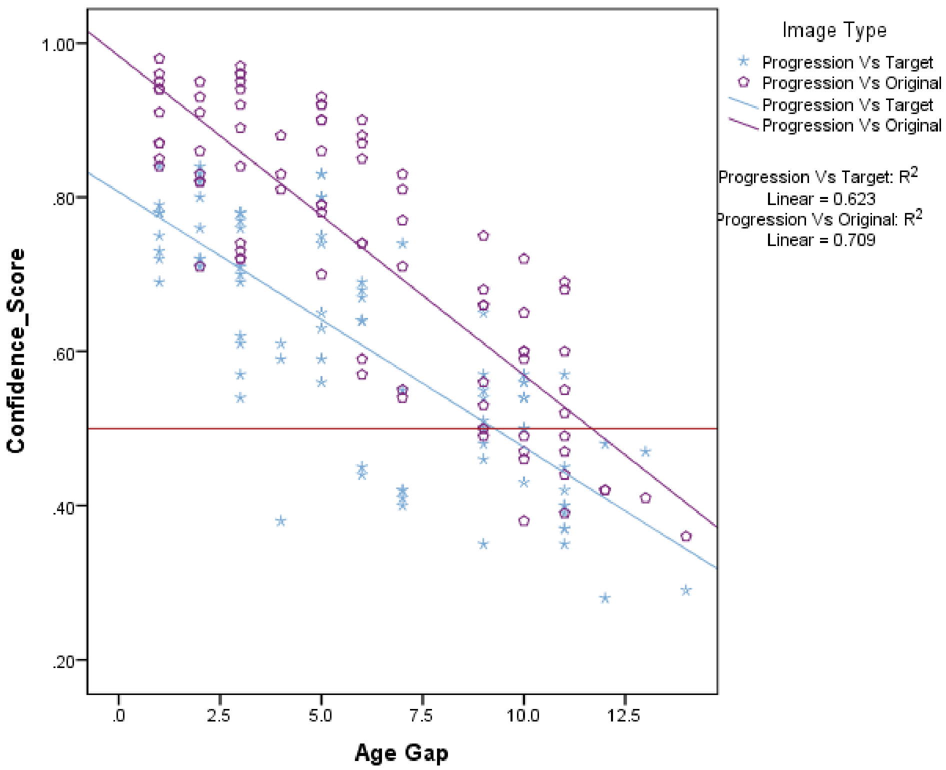


Figure 59: Face recognition rate in relation to age gap (years), comparing all progression types versus the target and the original image

Progression types were represented by Manual AP, Manual AP(Cr), KS(Cr) and KS(IA-Cr); KS(IA) was not included as previous analysis had shown a bias effect on the algorithm. With the Progression Vs Original showing an overall higher confidence score, this suggests the progression types were more similar to the original images [t(88) -5.031, p < 0.001]. Appendix 3 details the statistics report.

4.7.8 Overall

When the confidence scores for recognition of the manual age progressions were compared to scores of the original images subject M074 performed lower than other subjects (Figure 50.3). M074 had the lowest original age, meaning the age progressions were based on a face of a younger individual. All progression types are combined and visualised in Appendix 3.

5 Discussion

Key findings

- Experiment 1: The use of multiple images is beneficial for the facial identification of children
- Experiment 2A: The proposed age progression method for children recorded satisfactory levels of repeatability, with facial measurements at the Nasion (n) and Trichion (tr) showing the most inaccuracy
- Experiment 2B: Reduction of the resolution of an age progression image improves automated facial recognition for juvenile identification
- Experiment 2C: Manual age progressions are no more useful than the original image for facial identification of missing children

Key objectives 1-9 for Experiment 1 and 2 will be discussed in different sections. The realistic application of the methods used in this study will also be addressed in section 5.6 Human Vs Machine (Real case scenario application).

5.1 Discussion: Experiment 1

5.1.1 Benefits and limitations of multi-source database

These results support the prediction that using more than one image can lead to a higher recognition rate. Grouping faces together is not a new idea, Mu et al. (2014) have previously shown success across different FRS when more images of the same adult individual were used. This is the first time, however, that this has been tested on juvenile faces. Amongst the 4 databases used, the authors tested 11 images of 66 individuals from the FG-NET database. With the FG-NET containing juvenile faces, their results showed the lowest recognition rate when compared to other datasets. These other datasets were collected to address specific conditions, i.e. UMIST with a range of head pose; YALE with different lighting and facial expressions; ChokePoint simulating CCTV conditions. All databases except the FG-NET were adult subjects, the main interest of their paper was to demonstrate that performance could be improved with more training data, therefore, the authors did not explore the effect of facial growth or any other condition posed by the datasets. The design of Experiment 1 in the current study not only explored the difference between using one or five images, the study additionally explored if group tagging could improve juvenile recognition with age gap.

Face tagging, grouping or clustering is an effective tool to manage a large bank of photographs (Zhu et al., 2011) and this tool could be used to improve recognition of ‘faces in the wild’ when the subject will often provide more than one image. The main limitation of using multiple images is the ability to incorporate this idea into existing FRS for facial identification. The U.S. National Institute of Standards and Technology (NIST) had invited developers to assess the capability of different FRS in detecting and recognising unconstrained faces of children (Grother and Ngan, 2015). The institute had acknowledged the benefits of using multiple images per person; in order for developers to enter the test, the FRS must follow certain API requirements. This included a data structure to manage multiple images or video frames (Grother and Ngan, 2015; Grother et al., 2017, 2010). Developers were able to use multiple images of the same individual for all other tests, except 1:1 verification.

If the two faces are of different pose, lighting, resolution, and most importantly, age; what is the limit of such comparison using a single image? The results for single-source images (T1) in Experiment 1 were repeated twice, and the data was an average of the two different source images. Knowing that the difference between two images of the same age could generate a big difference in recognition rate, the test results could be biased towards the quality of the image set used. If multiple images of the same child are available, then their use would benefit face verification of a child at different ages with unconstrained image sets.

5.1.2 How subject age affects recognition rate when compared to other age images of the same person

Results suggest that faces younger than 1 year old are problematic for recognition as the source image. In the context of missing children, this suggests if the child went missing before age 1 year old, the available photographs are unlikely to be matched to a an older target. This supports the hypothesis that younger children are more difficult to discriminate.

Grother and Ngan (2014) suggest that "*Identification accuracy is strongly dependent on subject age*"; with a high false positive and false negative rate, algorithms are error-prone in the recognition of children's faces. However, examples showed that age estimation could be improved with a separate training dataset specifically for children (Antipov et al., 2016) and poor performance related to the lack of children's faces in the training dataset is a possibility. The recognition rate of a facial recognition software is directly related to the population of the training dataset. Most of these available datasets, such as Labelled Faces in the Wild (Huang et al., 2008) only consists of adults and therefore may not represent the recognition rate of children.

In experiment 1 the face of the child at approx. 5 years old was ranked as similar 80% of the time to faces up to 9 years of age. This suggests that the original out-dated images could potentially be matched to a target and be useful in machine-based recognition. However, this trend should not be generalised as the recognition pattern differed between subjects. With the increasing age gap, the face of FC001 achieved higher recognitions suggesting her face was more resistant to age-related changes in comparison to other subjects. Possible theories include:

- Some subjects could have a facial growth pattern that retains more juvenile features
- The variation of data between subjects could have been caused by image quality such as resolution, lighting, head pose, facial expression, and movement.
- The variation could also be also be gender specific with female faces changing less into adulthood than male faces.

5.1.3 <u>Identify the boundaries of the age gap for face recognition of children</u>

These results support the hypothesis that recognition rate drops with an increase in the age gap. The aim was to explore the maximum age gap over which face recognition is challenging. If the face of the child is no longer recognisable due to growth changes, when might age progression be useful? This section explores the relationship between age gap and age progression.

It is known that facial recognition is much harder to achieve for children and recognition rates will decrease with an increase in age difference (Lanitis, 2009) with significant problems associated with an age gap larger than 4 years (Ling et al., 2010). This study produced a mean recognition above 80% when the age gap was +2-4 years from 5 years old [FC001: Age>3 +2-6 years, FC002: Age >4 +2-4 years, MC001: >7 +2-4 years]. This trend will differ between individuals, as a child gets older and the face becomes more stable and older children could be recognised by FRS with a higher age separation. NCMEC (2016) recommends updating the age progression every 2 years, and every 5 years after 18 years old and this suggestion is supported by the results from Experiment 1.

By looking at the average of the three subjects studied in Experiment 1, the face of a child will most likely (80%) be recognisable within +2 years from around age 5. In addition, if the child went missing under the age of 5, an age progression with an age separation of +2 years would be less effective. Farkas and Hreczko (1994) suggest that the period of early rapid growth (growth spurt) of the head, face, orbits, nose, lips and mouth in both sexes was observed between 1 and 6 years of age. This could suggest why the result showed faces under the age of 5 are a lot less distinctive and less likely to be matched to an older age group.

The distribution of recognition in relation to the age gap was more gradual with FC002 when compared to FC001 and MC001. This could have been caused by the difference in video quality, the range of facial expression and the consistency of lighting and conditions. The difference could also have been caused by genetics as FC001 and MC001 are siblings with a degree of facial similarities or even a similar pattern of facial growth. If the variance is caused by the difference in growth pattern between each individual, this would suggest the recognition rate will vary between each child at different ages.

5.1.4 Dataset evaluation

The dataset used in Experiment 1 was static images extracted from video clips (Ani Acopian, 2014; Hofmeester, 2015, 2014). With changes in lighting, head pose and facial expression as the child grew, these images were relatively standardised where the subject-to-camera distance and the position of the face remained similar throughout the years. The pattern of recognition was different between the three subjects (FC001, FC002 and MC001), this may be due to:

- The footage of FC002 was made from static images only.
- Number of images varied, as it was dependent on the length of each video, therefore the number of images used for comparison was different between individuals.
- Quality of the images differed (e.g. the lighting of the videos was controlled but not standardised).
- These images were likely to have been affected by the video format compression with a loss in facial details; the image quality of FC002 was significantly lower in comparison to MC001 and FC001 (Table 32).
- The data of all three subjects were documented by their parents and later put onto the internet either by their parents or the subject themselves; the quality of videography could differ with the technological advances over the years and between the type equipment used.

Table 32: Image quality comparison of Experiment 1

		
FC001 (1080p)	FC002 (240p)	MC001 (1080p)
No copyright infringement intended under fair dealing for research, governed by Section 29 and 30 of the Copyright, designs and patents Act 1988.		

The biggest limitation for this part of the study was the number of longitudinal data available for comparison. With only 2F and 1M over 10+ years of growth from birth, the available data was not sufficient to suggest sex-linked differences. With the limited availability of longitudinal facial growth data, standardisation is another limitation. Facial recognition 'in the wild' will never be standardised, especially in the context of indecent images of children. The images obtained from these videos clips are already more standardised than images in the wild with controlled subject-to-camera distance, but it is unknown on how the variability of the dataset has an effect on the conclusion drawn.

The small sample size also limits the analysis between and within populations. Facial growth pattern will differ between faces, with so much noise in the data, and it is unknown how facial growth affected the recognition rate. Despite these differences, all three subjects showed similar recognition patterns in relation to the difference in the number of images and in age. However, since there were no distractors to enable the evaluation of false positive identification rates, as criticised by Grother et al. (2017), the results must be viewed with caution.

5.2 Experiment 2A

5.2.1 Evaluation of inter-observer error for the manual age progression method

A method of age progression was developed to guide practice. Forensic age progression is dominated by forensic artists, and a good knowledge of facial growth is necessary (Mullins, 2012; Taylor, 2000). The trends of facial growth have been described, with only a few examples of manual depictions quantifying growth by measurements (Farkas et al.,1994; Machado et al., 2017). Anthropometric measurements are quantifiable, therefore the artistic variability should reduce with the aid of the guided method for digital manual age progression. Experiment 2A compared the depictions created by three different practitioners, and the results showed that facial measurements involving the landmarks Nasion (n) and Trichion (tr) were the most unreliable. However, the landmark nasion had been previously reported as one of the most reliable landmarks (Campomanes-Álvarez et al., 2015), and this contradicts the conclusion drawn in the current study. The nasion is defined as *"The midpoint on the soft tissue contour of the base of the nasal root at the level of the frontonasal suture"* (Campomanes-Álvarez et al., 2015). The subject measured in the inter-observer study by the three practitioners was 3 years old and the region of the nasal root was underdeveloped, flat and ill-defined. This could have explained the decreased in precision of landmark placement.

Zygomatic width (zy-zy) measurement was consistent between the three practitioners (difference of 1.07%). However, this measurement has been previously reported as one of the landmarks with the highest variation (Campomanes-Álvarez et al., 2015). In this study the position of zygion was at the contour of the face rather than at the edge of the zygoatic bone. This was in order to measure the width of the face and may explain why this measurement remained consistent with little variation.

Anthropometry is able to guide feature location in age progression to a certain degree, but the process of age progression remains variable with some artistic interpretation, especially at certain facial features, such as face shape and jawline where anthropometry is less useful. These results suggest artistic depiction and performance varies between practitioners.

The reliability of age progression images has been explored previously (Erickson et al., 2016; Lampinen et al., 2015) by comparing depictions of the same individual created by different artists. In these studies the assessment of reliability was based on human perception and the authors concluded that the depictions varied in resemblance to the target. The current study using Microsoft Face API also showed a difference in confidence score between the depictions of different practitioners. Erickson et al. (2016) suggested that experience could be a contributing factor to performance. However, Lampinen et al. (2015) found no correlation between resemblance to the target and experience/training. Both studies had suggested various strategies to combat artistic variability; Lampinen et al. (2015) showed an improved recognition rate by morphing together of all the depictions created by different artists, whereas Erickson et al. (2016) suggested artists work together with models of an age predictive algorithm.

5.3 Experiment 2B

5.3.1 How recognition is affected by age gap in relation to different ages and genders

In Experiment 2B, 80 original images and age progression depictions from 24 subjects were compared to 83 veridical images. The results suggest recognition decreases as age gap increases. However, it is interesting to find that the faces were still being verified as the same individual with an average age gap of around 9 years. This trend is supported by the results from Experiment 1. These results may not be representative of all FRSs.

Experiment 1 showed that the younger age group (Age <1) had a lower recognition rate when compared to any other age. However, in Experiment 2, the confidence score difference between the original and the age progression images showed no correlation to the increasing age gap. This suggests that any differences between the original and the age progression images must be related to another factor.

Experiment 2B used Microsoft API to compare single images of the same individual at different ages, and facial similarities were given a score. The averaged confidence score of a facial image being recognised with an age gap of around 9 years was at 0.5. This is a weak similarity correlation. Therefore, as the age gap increases, the probability of false positive and false negative recognition will be higher. Experiment 1 used Google Picasa to match a group of faces against a large pool of longitudinal images of the same identity over time. By looking at age groups where 80% of faces were recognised, this could suggest a lower probability of false positives and negatives. The relationship between the age gap and false identification would require further testing.

There was no clear difference between male and female subjects in recognition rate and this could suggest growth related changes have a similar effect on FRS. Farkas and Hreczko (1994) indicated facial growth difference between the sexes, most noticeably in maturation age and in the periods of late accelerated growth. These changes were more significant in males from around age 10 years, including areas such as the width of the mandible, nose height, nasal tip protrusion and lip height (Farkas and Hreczko, 1994); most likely influenced by the hormonal differences during puberty. However, these changes were not reflected in the results, which suggest that even with facial changes both males and females were recognisable at a similar level with the increasing age gap.

5.3.2 How different conditions can affect machine-based recognition of age progressions

Age progression images create a likeness of the individual and it is not possible for these images to be exactly the same as the veridical image. To explore if the similarity scores could be increased by reducing the errors of the depiction, five different conditions were applied to reduce the information in the depictions: i.e. black and white, cropped, blur, resolution reduction, and combined. The confidence score of all conditions was compared to the age progressions.

Computer scientists have tested how different conditions can affect Deep Convolutional Neural Network (DCNN) based algorithms. Dodge and Karam (2016) tested four image classification algorithms and suggests that all models are sensitive to blur. Similarly, Grm et al. (2017) tested face verification performance on four DCNN based FRS; their study indicated that all models were sensitive to noise, blur, missing data and brightness (Missing data, in this case, were similar to partial occlusion of the face). Both studies found that contrast and compression were of low impact (Grm et al., 2017; Karahan et al., 2016). These study designs analyse how different conditions decrease the performance of an algorithm and some studies are designed to find the threshold of certain conditions (Boom et al., 2006; Lemieux and Parizeau, 2002; Marciniak et al., 2015; Wang et al., 2004). In order to explore if such conditions are able to improve the recognition rate of depictions, the changes in condition in this current study, such as blurring and resolution reduction, were not as extreme. The aim was not to find a threshold, but to analyse if a change in condition can improve recognition.

The performance of a face recognition algorithm can remain similar with images across different resolutions (Boom et al., 2006; Lemieux and Parizeau, 2002; Marciniak et al., 2015; Wang et al., 2004). With resolution reduction, the current study suggested a significantly higher recognition for both male and female subjects when compared to other conditions and the original image. The improved recognition suggests conditions such as resolution reduction and blurring decrease the differences between the two images and make the depictions more recognisable.

Image processing can have a negative effect on recognition (FISWG, 2016). With previous research suggesting that colour information does not have a significant effect on performance (Grm et al., 2017; Karahan et al., 2016), it was surprising to discover the black and white condition performed the worse. Although concealing what is unknown could be beneficial in some circumstances (Abudarham and Yovel, 2016; Erickson et al., 2016; Lampinen et al., 2015), the cropped condition had a negative effect on recognition. However, both human and machine-based recognition suggests a positive effect for blurring and resolution reduction.

One suggestion for improvement to FRSs is often related to the training database (Buolamwini and Gebru, 2018): for example, increase the number of low-quality images (Dodge and Karam, 2016) or more profile images (Mehdipour Ghazi and Kemal Ekenel, 2016) during training. Perhaps the low recognition rate is related to the lack of certain conditions within the training set; for example, the lack of black and white conditions could lead to a lower recognition rate of these images, and same for the cropped images. However, the effect of cropping (i.e. the percentage of the face shown in an image) could be a factor in how the algorithm recognises the face as a face. Increasing the number of low-quality images in the training dataset for the algorithm may have an effect on the performance of high-quality images (Dodge and Karam, 2016), and this could increase of the number of false positive and false negative recognitions, which may not be beneficial to an FRS.

Practitioners could test what imagery works best with their FRS by creating different conditions of the same image. Image manipulation can be produced on age progression depictions or out-dated photographs and be compared to a photograph at an older age. Figure 60 used an age progression of subject FM035 as an example.

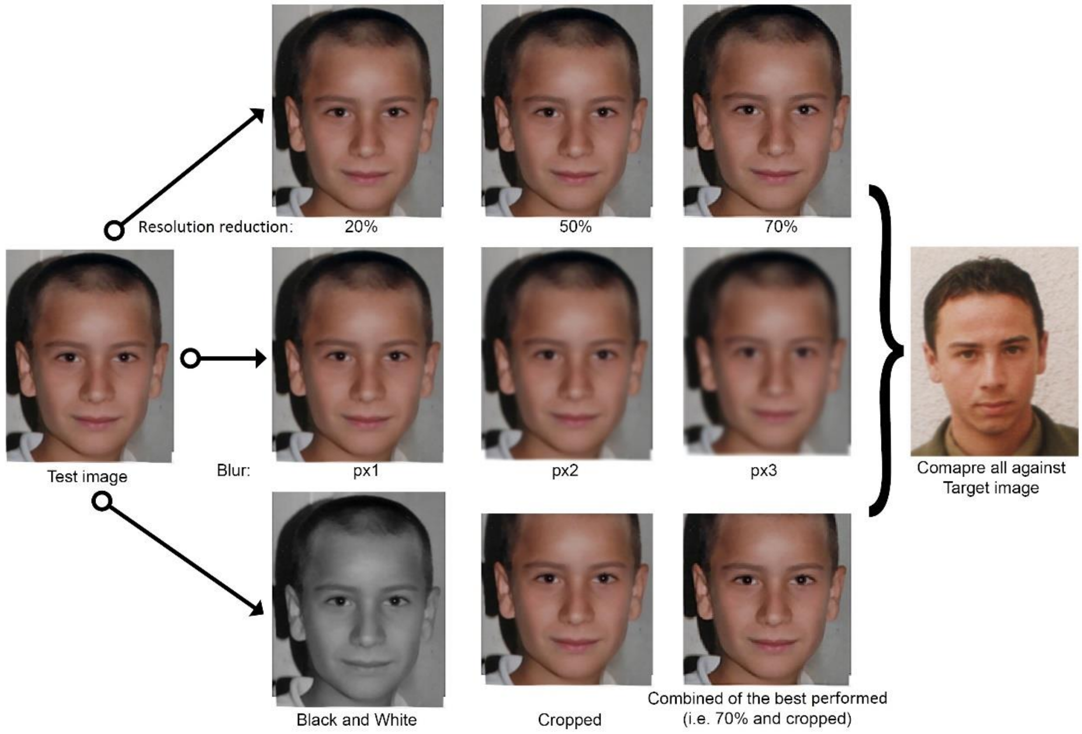


Figure 60: Workflow for image testing

5.3.3 Performance of the age progression image as compared to the original image

Although the trend for age progression recognition was slightly higher than for the original image, the difference was not statistically significant. This suggests that the original out-dated images could be just as useful as an age progression image in a forensic case. However, neither performance was comparable to a recognition. The trends for the original image and the age progression image were very similar, where recognition degrades around an average age gap of 9 years.

Previous research addressing human recognition also suggested no significant difference between out-dated images and age progressions (Lampinen et al., 2012a). This is perhaps because even though both image types are different from the target image they remain somewhat similar when compared. The higher recognition rate (58%) for the age progression than for the original could indicate that the growth prediction was similar for some individuals. The data used to develop this method was population specific, which could have been a contributing factor, especially when the FG-NET database is population unspecific.

With results suggesting no significant differences between using out-dated and age-progressed images, all possibilities should be included when using machine-based recognition, especially as Experiment 1 has shown advantages to using more images. Most users who deal with the identification of indecent images of children (i.e. law enforcement) will most likely be using a commercial black box system, meaning they do not know exactly know what images the algorithm works best with or what facial features the system is identifying.

5.3.4 Dataset evaluation

Images within the FG-NET dataset are often variable in quality and age gap. Some images were black and white or decolourised due to the age of the photograph, and it is unknown when some of these images were taken. This means some of the conditions for black and white comparisons (12/83; 14.5%) were missing. Some were already cropped, where the outline of the full head differed between images. Therefore the variability of these images could have affected the difference in conditions.

The experiment for image variability has shown that a different target image at the same age was able to generate different recognition rates, and at times, was able to generate a positive match. This would no doubt have an effect when images were compared for their similarities. It is a difficult factor to control, especially when considering that indecent images of children will never be standardised.

5.4 Experiment 2C

5.4.1 Comparison of the recognition rate between manual and machine-based age progression methods

Machine-based age progression methods differ between previous studies as the stimulation of 'growth' varies with the design and training dataset. The study by Kemelmacher-Shlizerman et al. (2014) provided sufficient comparable images, i.e. from the FG-NET database, to represent a machine-based method. The ideal situation would be to obtain different algorithms to produce such imagery; however, with limited research in this field, studies who have used comparable data (i.e. FG-NET) have only provided a small number of examples (Liang et al., 2007; Ramanathan and Chellappa, 2006). Experiment 2C offered a framework for comparing different methods of age progressions.

The manual age progression showed no significant difference when compared to the original out-dated photograph; this is similar to the results from Experiment 2B. When the performance of the blended age progression KS(IA) from Kemelmacher-Shlizerman et al. (2014) was compared to the manual age progression, KS(IA) performed significantly better in comparison to the manual age progression and the out-dated image. However, the images of KS(IA) were utilised with the same external background as the veridical images and these results suggest that this created a falsely high recognition rate. Therefore, a comparison with the cropped versions, i.e. KS(IA-Cr) was more representative of the KS age progression method; these results suggest the manual method was similar to the machine-based method with no statistically significant difference.

If the age-progression of the missing child was blended onto a template image of a child at a similar age, this template could be used to provide the head-shape, face-shape, skin-tone and hair-style. With the background excluded, the internal faces of the KS(IA-Cr) condition also had the same illumination as the veridical images. KS(IA-Cr) performed slightly better than the non-illuminated condition KS(Cr), but this is not surprising as the correct texture will benefit recognition. This is not, however, representative of a real case scenario where the veridical image is unknown.

It is arguable that the KS(IA) blended images from Kemelmacher-Shlizerman et al. (2014) were not designed for such a FRS comparison. The authors are aware of such similarity bias and have assessed human recognition using different images of the same individual at the closest age. With 8,916 comparisons, their results indicated that the original image (44%) and the machine-based age progression image (37%) achieved similar recognition rates. This could indicate their results were similar to this study, with no significant difference between using the out-dated or age-progressed image.

Some machine-based age progression studies have assessed the accuracy of the models through visual human assessment (Gibson et al., 2009; Liang et al., 2007; Scherbaum et al., 2007, 2007; Shen et al., 2014). Some used quantitative methods such as shell deviation colour maps (Koudelová et al., 2015; Matthews et al., 2018); some used distance metrics (Bukar and Ugail, 2017; Lanitis et al., 2002); and some designed experiments based on human recognition (Bukar and Ugail, 2017; Kemelmacher-Shlizerman et al., 2014). The current study tested the images using ‘off-the-shelf’ FRS, which could potentially be more applicable for realistic comparisons, i.e. analysis in aid of law enforcement. Ramanathan and Chellappa (2006) used an FRS to quantify the accuracy of their model, which is perhaps the most comparable methodology. Their study compared the original and the age-progressed image(s) to a gallery from the database (233 images, n=109) containing the target image; the FRS was able to make rank based suggestions, and the age-progressed condition showed a better recognition rate than the original image. The authors suggest that the low rank-1 recognition (15%) is affected by the difference in image quality such as illumination and head pose.

The FRS Ramanathan and Chellappa (2006) used was an early eigenface approach developed by Turk and Pentland (1991). Although this method is useful for reducing the dimensionality of the original data, it had been described to have poor discriminating powers (Bhele and Mankar, 2012). With the advancement in FRS over the past decade, using a more recent FRS is more relevant.

The resemblance to the target was compared between the manual method and the machine based method: results suggest as age gap increases, the two progression types become more dissimilar to the original image and also to each other. However, when the confidence score was compared, both the progressions were more similar to the original image than to the target. After all, the progressions were based on the original image.

5.4.2 The effectiveness of the proposed manual age progression method

With the manual age progression condition showing no significant difference when compared to the original photo or the machine-based technique KS(IA-Cr), this suggests all methods are similar. It is arguable that with no significant benefits, a method of age progression is not very effective; however, with the vast amount of constraints and variability, the methods do not harm an investigation and may even benefit it as using multiple images is shown to increase the probability of recognition.

Machine-based age-progression systems are developed using face databases of children at different ages, and this can be longitudinal or cross-sectional. However, age progression depictions are not often accurate enough to support forensic investigations (Lanitis and Tsapatsoulis, 2016) and machine-based methods are not used by the main centre in locating missing children (NCMEC, 2016). This suggests that machine-based methods are not recognised to be an effective age progression tool by practitioners, and this was reflected in the survey conducted by Erickson et al. (2016). The authors interviewed eight forensic artists; only one practitioner used computer algorithms, and two used growth norm database. This could suggest quantifiable measurements are not favoured by the artists. Perhaps the 'accuracy' of an age progression is influenced by the visual expertise of an artist (Erickson et al., 2016). The practitioner who used a computer algorithm also used other editing tools, and this indicates the algorithm did not function alone and required a certain level of 'artistic' interpretation.

With their superior visual experience, artists are able to generate better imagery than those of non-artists (Kozbelt, 2001; Vogt and Magnussen, 2007). Manual age progressions carried out by artists are often photorealistic, and the realism of a face perhaps influences human perception on the 'accuracy' of the depiction. However, if the recognition task is carried out by a computer, how 'photorealistic' does the image need to be? An image does not have to be aesthetically pleasing for an FRS (FISWG, 2016), and research suggests that pre-processing an image before an FRS can benefit recognition (Fookes et al., 2012; Mehdipour Ghazi and Kemal Ekenel, 2016). However, image pre-processing may affect FRS differently, and in some cases may degrade performance (FISWG, 2016). This was also shown in the current study when the different conditions were applied.

Age progression depictions vary between artists, and experience could be a factor (Erickson et al., 2016). However, some research suggests otherwise (Lampinen et al., 2015). Not all artists interviewed in the two studies (Erickson et al., 2016; Lampinen et al., 2015) were specialised in child age progression and most were focused on adults; however, their results showed no correlation. Some artists performed better in comparison to others (Lampinen et al., 2015), it is therefore difficult to assess the accuracy of manual age progression in general. Erickson et al. (2016) suggest "*The best way to increase reliability would be to develop techniques that create better likenesses*." With no current standard on how an age progression should be carried out, the current study introduced a quantitative method in extrapolating measurements based on anthropometric research. The idea provides a more guided method to manual age progression, thus reducing the level of artistic variability between practitioners. However, with the inter-observer study showing a variation in landmark measurements, a level of variability in this method will remain along with the difference in artistic interpretation.

For an age progression to be effective, the images would have to represent current appearance and be able to provide improvements over an outdated photograph (Lampinen et al., 2010). Understanding the changes in facial growth from childhood to adulthood is key to developing or using any age progression methods or tools (Taylor, 2000).

5.5 Critical review

5.5.1 The methodology

This research explores how well an FRS is able to recognise a child's faces over periods of facial growth and only considered face verification. The addition of distractors to evaluate false positives/negatives will strengthen future studies. Similar work has been carried out by the NIST using a database of child exploitation images (Grother and Ngan, 2015). The use of such a database is ethically challenging, and whilst academics can use open-source databases such as MegaFace (100M faces from Flickr) by Miller et al. (2015), it contains a combination of all ages. The addition of distractors (random unmatched faces) can be collected from the internet. For example, Antipov et al. (2016) collected a private dataset of 5,723 images of children between 0-12 years old from the internet. However, the legal, ethical and reliability issues of such internet image collection should be considered.

The recognition rate in comparing unconstrained facial images will most likely be influenced by the quality of the original (input) image, with factors such as head pose, facial expression, lighting, distortion, and resolution. These factors will also limit the success of an age progression. The quality of the 'ground-truth' (target) image will also be a significant factor affecting the recognition rate. Differences in the quality between the source and the target images are unavoidable, especially when the photographs were taken using different equipment under different conditions. Manual age progression is subjective and can vary between different practitioners; this is also a limiting factor when the progression images were only produced by one researcher.

In Experiment 2C, machine-based age progression was represented by only one algorithm with a limited number of comparable examples derived from Kemelmacher-Shlizerman et al. (2014). It would be ideal if their algorithm was obtained for comparisons with other datasets.

5.5.2 <u>The algorithms</u>

Due to limitations in the FRS, the video clips in Experiment 1 had to be translated into static images, as the algorithm was unable to perform facial recognition on video content. Indecent images of children are often presented as video imagery, and this presents even more challenges for an FRS including false detections with an absence of a face in the sequence, multiple subjects, varying solution/scale and head pose, and the loss of facial details from motion blur and format compression (Grother et al., 2017).

Microsoft API is able to provide a confidence score between two images for comparison. This was used in an attempt to quantify and compare the resemblance between two facial images; but the experimental format was 1-to-1 verification, which will differ to forensic scenarios, where it will most likely be 1-to-many identification. Therefore, the confidence score in this study is not representative of the confidence level in an identification scenario. With the growth related changes to the child's face, future research should consider manipulating the acceptance level, where a target face will be more likely to appear in the pool of possible matches.

Google Picasa is a commercial software and therefore the algorithm is a black box system, but additional tagging of the same face at different poses did not generate new recognition, rather it is a collation of 'hits' based on the tagged images. Therefore are algorithms able to learn from the tagged faces and produce a new form of recognition in a manner similar to the idea of 3D alignment proposed by Taigman and his colleagues (2014)? Not just two-dimensionally, but by collating information from different head poses to create a three-dimensional template?

Google Picasa had been useful in analysing the dataset presented in Experiment 1; however, Google recently discontinued the support for Picasa and it is no longer available unless you have previously obtained a copy of the software. With the constant improvements, it is likely that different FRS will present different results (Buolamwini and Gebru, 2018), the current study has shown how an FRS can be tested for application in the facial recognition of children.

5.5.3 The age progression method

Most juvenile age-progression literature uses FG-NET, a longitudinal dataset with unconstrained images of individuals across different ages. It was quite difficult to blind-test using FG-NET, as this database does not contain information about ethnicity or hereditary factors (images from relatives) and this will affect the quality of the age progression. It was especially difficult to depict the accurate skin tone of the individual based on the very young photographs (under 3 years of age). The technique of age progression uses images of other children of a similar age; therefore, the age progression will be skewed towards the appearance of the reference images used. It is also worth noting that the growth guidelines (Farkas and Bolton standard) were developed on North American children. These are population specific, but the face and head shape within and between populations are more diverse than this single template. The method described in the current study could reduce artistic variability and is more tailored to each image/individual; however, the accuracy of the method could be improved by using measurements from specific populations or similar face shape groups, where the growth pattern from family members could be useful in predicting the facial growth trends.

Most FRS reduces in recognition rate when non-frontal face images are used (FISWG, 2016). Indecent images of children will most likely be images with varying head poses; as recommended by FISWIG, the recognition rate of such images could be improved by the application of pose correction (FISWG, 2016; Mehdipour Ghazi and Kemal Ekenel, 2016). This procedure could be useful in correcting the original image before conducting an age progression, as frontal images would benefit the anthropometric measurements on an image. These measurements will also be affected by the subject-to-camera distortion which could also potentially be adjusted (Stephan, 2015).

5.6 Human Vs Machine (Real case scenario application)

Real-world forensic cases are discussed below to address how a machine differs from human recognition.

5.6.1 The application of FRS in forensics

With Deep Convolutional Neural Network (DCNN) being considered as the forefront of face recognition (Phillips et al., 2018; Ranjan et al., 2017; Rawat and Wang, 2017). Grother et al. (2017) criticised the insufficiency of the open universe in academic research with a lack of measures in false positive identification rates. The authors addressed the consequences of false positives in forensic settings and questioned the ability of academic research related to DCNN to be used "off-the-shelf". The distinction between open- and close-universe face identification is the ability to provide a confidence of the identity predicted (Becker and Ortiz, 2013). Most of the FRS in academia have been tested on existing standard benchmarks and datasets; Becker and Ortiz (2013) have commented on the lack of consideration in the realistic application of such databases; the authors chose four face identification platforms (i.e. research, client-side, consumer and cloud-based), and selected a few algorithms to represent each group. The performance of the different FRS was compared under an open-universe setting, Becker and Ortiz (2013) concluded that research algorithms are able to show a high performance with the greatest flexibility, but they are often difficult to set up.

In recent years, certain DCNN algorithms have reported an accuracy above 99% when tested against the same database (Rawat and Wang, 2017). Indeed, the forensic standards for false positive may not correlate to this accuracy. However, the meaning of such accuracy should not be undermined, rather, should be further tested. In relation to missing children, unconstrained imagery could be similar to those presented on social media. If the performance of an algorithm is affected by the training set, should the choice of algorithm be dependent on the environmental similarity of the imagery? When the face of a child is not a permanent feature, the number of matches could be increased by lowering the benchmark for a match. This will no doubt increases the false positive and false negative rates, however, if this could increase the chances of the child being matched, should this be considered?

The reported accuracy of an FRS should also be interpreted with care, as suggested by Buolamwini and Gebru (2018), face recognition algorithms could vary in error rate in response to different gender and skin-type. Algorithmic discrimination is related to skewed populations of the training dataset (Bolukbasi et al., 2016; Caliskan et al., 2017), and the benchmark datasets such as the LFW had also been criticised (Buolamwini and Gebru, 2018). Although these biases described are related to gender classification and labelling, the performance of face recognition could also be skewed by the difference in gender and ethnicity. Buolamwini and Gebru (2018) supports algorithmic transparency and accountability; the authors encourage developers to report on the phenotypic and demographic representation with the face datasets.

5.6.2 Machine and human performance in forensic cases of juvenile age progression

Manual age progression work is subjected to high levels of an artistic impression; images generated can vary between practitioners and will most likely differ to the "ground-truth". The likeness between the progression and the veridical may be fundamental to the identification of a missing child.

Using four examples of forensic cases, the differences between an FRS and human recognition is discussed below. The 'reliability' of these four cases had been widely discussed by the general public on social media forums (LIAM, 2015). With the complexity of such investigations, it is uncertain how much, if any, the age progression image aided the identification of the individual (CBC News, 2002; Goldman, 2009; Malarek, 2002; Whitefoot, 2018). There are also examples where individuals recognised themselves through an age progression (Melendez, 2012; Wagner, 2015).

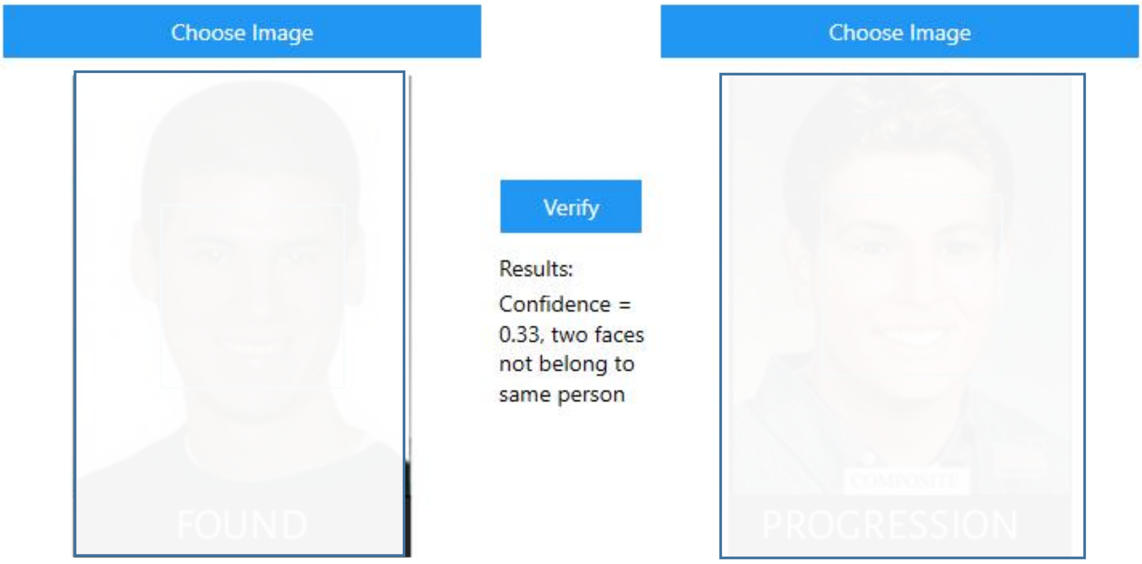
In cognitive psychology, there are different types of facial recognition. Using the public to comment on the likeness between two images involves unfamiliar facial recognition, and recognising a face you know, or even recognising your own face is familiar face recognition. Humans recognise faces differently according to the distinctiveness and familiarity (Johnston and Edmonds, 2009), and that could have a contributing factor on how we perceive 'resemblance' of faces, such as age progressions.

Below are examples of four forensic cases from LIAM (2015); these individuals went missing at a young age and were later found (Figure 61). LIAM (2015) carried out an online survey for the general public to vote for the "best progression". Only 6% of the participants indicated 'They are all bad!' with all four progressions receiving between 21-27% of votes to be the 'Best progression'. This suggests the majority of the participants thought the age progression was somewhat similar to the image of the individual when found.

The participants already know the age progression is a depiction of the match image, the comparison of recognition here is not comparable to a real case scenario. However, the dataset here is valuable, as these comparisons are rare; individuals who are later found may not want to be identified for their images to be used in such a way. These cases are useful in assessing the practice and reliability of age progression.

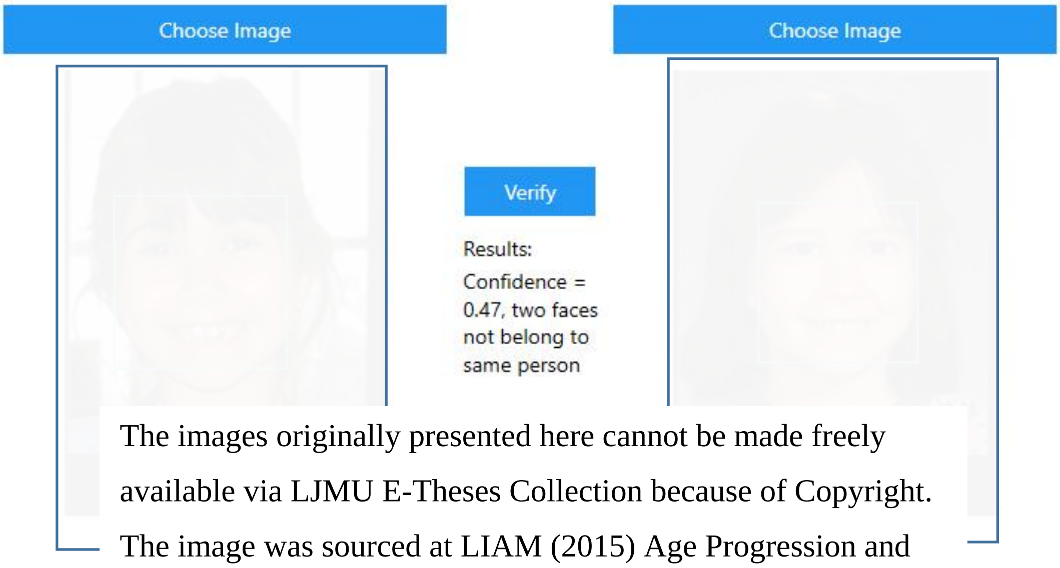
Using Microsoft Face API, subject 1 and 3 in Figure 61 presented a lower confidence score when the two faces were compared. The two depictions (subjects 1 and 3) had an age gap of 18 and 22 years, as suggested by previous literature and results from the current study, the low confidence score is likely to be related to the age gap. Only subject 4 achieved a positive match between the depiction and target image; it is also the only image with a different head pose and ethnicity. It is uncertain if these differences could be a contributing factor.

Subject 1: Missing at 6 weeks old and found 22 years later

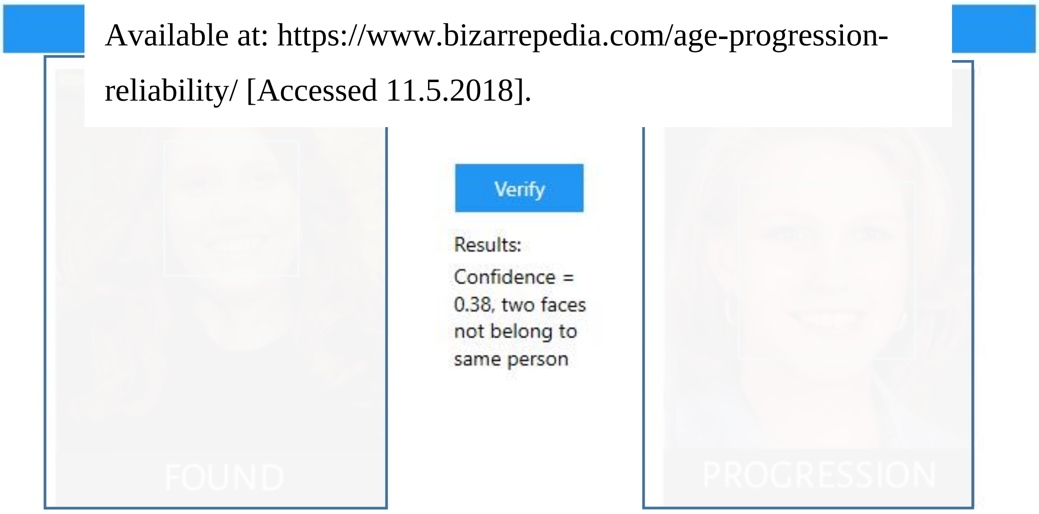


Age Gap: 22 year progression

Subject 2: Missing at 3 years old and found 5 years later

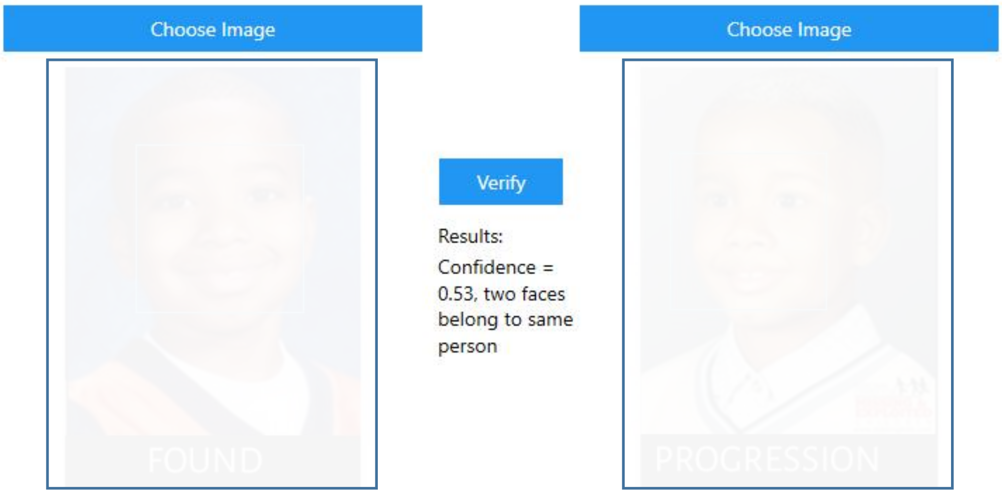


The images originally presented here cannot be made freely available via LJMU E-Theses Collection because of Copyright. The image was sourced at LIAM (2015) Age Progression and It's Reliability. Age Progression and It's Reliability, [online] Available at: https://www.bizarrepedia.com/age-progression-reliability/ [Accessed 11.5.2018].



Age Gap: 18-year progression

Subject 4: Missing at 3 years old and found 6 years later



Age Gap: 6-year progression

Figure 61: Forensic age progression cases

The age progression of subject 4 appeared to be the most unrealistic face amongst the four depictions, however, the effect on recognition was not reflected in the short survey by LIAM (2015). The “uncanny valley” evokes a negative reaction, and such perceptual tension could change the observer’s behaviour (MacDorman and Ishiguro, 2006; Moore, 2012; Seyama and Nagayama, 2007). It is uncertain if these negative reactions will affect the ability of humans to recognise faces, however, with the FRS showing a match, perhaps the uncanny valley of a face is less significant to a machine, where the image does not have to be aesthetically pleasing, as suggested by FISWG (2016).

5.6.3 Quantifying the ‘Reliability’ of juvenile age progression

To generate a more accurate progression, most of these cases in Figure 61 would have needed parental or relative references; hereditary facial features could have contributed to the success rate of these age progressions. Microsoft Face API was unable to recognise all individuals in Figure 61 as the same individual, except Subject 4. These comparisons suggest artists’ manual depictions deemed similar by human recognition, were not deemed similar by Microsoft Face API. It is difficult to quantify the likeness of age progression with human recognition; psychologists often test the recognition using array images mixing the test image within a pool of faces (Charman and Carol, 2012; Lampinen et al., 2012b, 2012a) or in some studies suggesting the likeness using a Likert scale (Erickson et al., 2016).

For the computer scientist, the accuracy of age progression depictions has been quantified using the following different methods: classification by ranking age progressions against a reference dataset (Lanitis and Tsapatsoulis, 2016); a score describing the similarity between an age progression to the original and the veridical image (Bukar and Ugail, 2017; Lanitis et al., 2002); face matching study for participants to choose a more convincing likeness between an older age and the age progression against a younger age (Kemelmacher-Shlizerman et al., 2014); participants were asked to rank the different age progressions by likeness when compared to the original photograph (Bukar and Ugail, 2017); visualising the surface differences between the 3D age progression model and the individual with 3D data using shell deviation colour maps (Koudelová et al., 2015; Matthews et al., 2018).

Previous studies are divided into three main categories to quantify the 'reliability' of an age progression:

1 Human recognition
2 Machine-based recognition
3 Shape difference comparison

The different experimental set up of human face recognition can generate different results, and psychologists have considered the sensitivity of memory in facial recognition (Bruce et al., 1991; Lampinen et al., 2012a). Age progressions may bear a resemblance to the target image, but memory is sensitive to the configurations of facial features (Bruce et al., 1991). Therefore, under what circumstance could these depictions be recognised? What is deemed similar in human recognition could be different for a machine-based facial recognition system. The shape is important to indicate accuracy between two objects, but the difference in shape is not the only factor that affects recognition. Therefore how much of a difference is needed to have an effect on recognition?

Face verification is a method of identity authentication to determine if two faces are similar. This raises another question on the effectiveness of using FRS to compare 'resemblance' faces. If humans suggest those forensic cases shown in Figure 61 are somewhat similar, should the confidence level be lowered to match what humans deemed similar? If 'resemblance' faces can trigger human face recognition, can computers do the same? Does it matter when the machines fail to see the similarities that humans perceive? Research in DCNN is most likely trained using large unconstrained datasets, and this increases the difficulty in analysing the source of errors and problems (Grm et al., 2017). This suggests it is unclear what facial similarities mean to a machine; in order to address how an algorithm can improve the recognition rate of 'resemblance' faces, it is important to research into facial feature sensitivity of FRS by comparing how different parts of the face can affect recognition rate, similar to the work of Grm et al. (2017) where features were blanked and compared.

In forensic settings, most FRS are monitored by human operators who then make the final recognition decisions (Grother et al., 2017), and the combination of the two methods of recognition has been shown to offer the highest possible level of facial identification accuracy (Phillips et al., 2018). Perhaps an FRS can be useful in narrowing down the searches for the human operators?

6 Conclusion

Indecent images of children will, by their very nature, be unstandardised and therefore using a single-image for verification and identification will limit the recognition rate for images in the wild. Group tagging has been shown to increase verification rates and this could be a beneficial method for face recognition in the wild. This research recommends that researchers and practitioners use, or design their algorithm with the ability to group multiple images of the same individual. The methodology here is by no means representative of the accuracy of identification, but it does show how well an algorithm is able to recognise the same face across different ages. Therefore, using an FRS could be beneficial to increase the validity and objectivity of the study. However, we have to be aware that recognition provided by an algorithm is somewhat different from human recognition. There are cases where a missing child was found based on human recognition, but with the increasing displacements of populations and an overload of media information, human recognition may not be an effective method. Testing the feasibility of algorithms in the recognition of children's faces that are years apart could potentially save time and resources in the search for a missing child. The feasibility of FRS is especially important when dealing with indecent images of children to minimise the workload and deal with moral and stamina challenges related to human recognition.

Experiment 1 showed that FRS could recognise images of the same child within a 2 year age gap from around age 5, and group tagging could lead to higher recognition rates. Faces at age 1 and below are problematic for recognition at any other age.

An anthropometric based method for manual age progression was developed, where practitioners could have a more guided process of manual age progression. Experiment 2A suggests the measurements are able to guide the positioning of the features to a certain level. However, facial measurements involving the landmark Nasion (n) and Trichion (tr) showed inconsistency. Even with the help of the guided method, the process of age progression remains variable with artistic interpretation.

Experiment 2B and C suggest algorithms (Microsoft API) could recognise images with an age gap of around 9 years. With no difference in recognition between manual/machine-based age progression depictions and the original images, the out-dated images could be just as useful as an age progression and should be included when using machine-based recognition.

Different conditions can have an effect on machine-based recognition, and Experiment 2B suggests resolution reduction can have a positive effect, where the black and white and cropped conditions showed a negative effect.

6.1 Future research

This study has only considered face verification, the identification of a child using age-progressed images amongst many should be further tested. To have more comparable data, similar to the FG-NET, building a shareable database, such as the Open-access Child Aging database should be expanded to assist research in this field.

The validity of the guided method for age progression proposed in this study requires further testing. The variable landmarks used in this study, Nasion (n) and Trichion (tr), should be changed or defined to reduce inconsistency. Practitioners are encouraged to use data for the 11 facial anthropometric measurements from different face shapes or populations for comparison. However, this method is limited by constraints, such as head pose and camera distortion, and developing a correction process could optimise the process of age progression.

Research on the relationship between face shape and facial growth would help increase the accuracy of related methods. Growth studies are mostly population-based, with the diversity of the difference in face shapes, perhaps the difference in facial types could help develop a better prediction model for age progression.

In order to enable practitioners to tests other FRS under different conditions, the method of testing a black box FRS should be further expanded and standardised. Having the ability to establish which condition works best could potentially improve the probability of a match. However, most FRS are developed using adult data, and a child's face is not a permanent biometric and cannot be treated the same as an adult's face. Developing an FRS with focus on child growth, with the ability to account for the difference in age gap, could potentially improve facial recognition of children.

The acceptance rate in comparing 'resemblance' images such as an age progression depiction should be further discussed. It is near impossible to generate an exact likeness with external factors such as hairstyle, body modification, makeup, and other forms of alteration to the face. Should these images be treated the same as a normal facial comparison? Perhaps the tolerance of an FRS should be explored, it is unclear whether certain features are more superior for face recognition. By altering different facial features, this could provide a better understanding of the perception of faces by a machine.

If humans are able to recognise an individual based on depiction, perhaps an FRS could be trained to do similar tasks. This could increase false positive and negative identifications, however, if the purpose is to generate an investigative lead rather than an identification tool, the practicality of such tools should be explored.

7 References

Abudarham, N. and Yovel, G. (2016) *Reverse engineering the face space*. Journal of Vision, 16(3):40, pp.1-18. DOI:10.1167/16.3.40

Ani Acopian (2014) *Ani Time Lapse: birth to 14 years old in 2 minutes*. [online] Available at: https://www.youtube.com/watch?v=K_lPKPpZUtA [Accessed 25.4.2017].

Antipov, G., Baccouche, M., Berrani, S.-A. and Dugelay, J.-L. (2016) *Apparent age estimation from face images combining general and children-specialized deep learning models*, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, IEEE, pp.801-809. DOI:10.1109/CVPRW.2016.105

Avon, S.L. (2004) *Forensic odontology*. Journal-Canadian Dental Association, 70(7), pp.453–458.

Bachmann, T. (1991) *Identification of spatially quantised tachistoscopic images of faces*. European Journal of Cognitive Psychology, 3(1), pp.87–103.

Becker, B.C. and Ortiz, E.G. (2013) *Evaluating Open-Universe Face Identification on the Web*, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE, pp. 904–911. DOI:10.1109/CVPRW.2013.133

Bhele, S.G. and Mankar, V.H. (2012) *A review paper on face recognition techniques*. International Journal of Advanced Research in Computer Engineering and Technology (IJARCET), 1(8), pp.339-346.

Bhuvaneshwari, K.V., Abirami, A. and Sripriya, N. (2017) *Face Recognition Using PCA.* International Journal of Engineering and Computer Science, 6(4), pp.20820-20826. DOI:10.18535/ijecs/v6i4.04

Bishara, S.E., Peterson, L.C. and Bishara, E.C. (1984) *Changes in facial dimensions and relationships between the ages of 5 and 25 years*. American Journal of Orthodontics, 85(3), pp.238-252. DOI:10.1016/0002-9416(84)90063-0

Bjork, A. (1963) *Variations in the growth pattern of the human mandible*. Journal of Dental Research, 42(1):2, pp.400-411.

Blanchette, M.E., Nanda, R.S., Currier, G.F., Ghosh, J. and Nanda, S.K. (1996) *A longitudinal cephalometric study of the soft tissue profile of short- and long-face syndromes from 7 to 17 years*. American Journal of Orthodontics and Dentofacial Orthopedics, 109(2), pp.116-131.

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V. and Kalai, A. (2016) *Man is to Computer Programmer as Woman is to Homemaker?*, in: ArXiv:1607.06520 [Cs, Stat], Presented at the 30th Conference on Neural Information Processing Systems, Barcelona, Spain, pp.1-9.

Boom, B.J., Beumer, G.M., Spreeuwers, L.J. and Veldhuis, R.N.J. (2006) *The effect of image resolution on the performance of a face recognition system*, in: Proceedings of the Ninth International Conference on Control, Automation, Robotics and Vision (ICARCV). Presented at the 9th International Conference on Control, Automation, Robotics and Vision, ICARCV '06, Piscataway, NJ, USA: IEEE.

Bourlai, T. (Ed.) (2016) *Face Recognition Across the Imaging Spectrum*. Switzerland: Springer International Publishing.

Broadbent, B.H. and Broadbent, B.H., Golden, W.H. (1975) *Bolton standards of dentofacial developmental growth*. Saint Louis: Mosby.

Bruce, V., Doyle, T., Dench, N. and Burton, M. (1991) *Remembering facial configurations*. Cognition, 38(2), pp.109-144. DOI:10.1016/0010-0277(91)90049-A

Bruce, V., Healey, P., Burton, M., Doyle, T., Coombes, A. and Linney, A. (2013) *Recognising facial surfaces*. Perception, 42(11), pp.1200–1214.

Bukar, A.M. and Ugail, H. (2017) *Facial Age Synthesis Using Sparse Partial Least Squares (The Case of Ben Needham)*. Journal of Forensic Sciences, 62(5), pp.1205-1212. DOI:10.1111/1556-4029.13523

B Bukar A.M., Ugail H., Hussain N. (2017) On Facial Age Progression Based on Modified Active Appearance Models with Face Texture. In: Angelov P., Gegov A., Jayne C., Shen Q. (eds) Advances in Computational Intelligence Systems. Advances in Intelligent Systems and Computing, vol 513. Springer, Cham.

Bulut, O. (2010) *Blind study of age progression evaluation*. MSc Thesis, The University of Dundee.

Buolamwini, J. and Gebru, T. (2018) *Gender shades*, in: Conference on Fairness, Accountability and Transparency, pp. 77–91.

Burton, A.M., Schweinberger, S.R., Jenkins, R. and Kaufmann, J.M. (2015) *Arguments Against a Configural Processing Account of Familiar Face Recognition*. Perspectives on Psychological Science, 10(4), pp.482-496. DOI:10.1177/1745691615583129

Cakan, D.G., Ulkur, F. and Taner, T. (Uğur) (2012) *The genetic basis of facial skeletal characteristics and its relation with orthodontics*. European Journal of Dentistry, 6(3), pp.340-345.

Caliskan, A., Bryson, J.J. and Narayanan, A. (2017) *Semantics derived automatically from language corpora contain human-like biases*. Science, 356(6334), pp.183–186.

Campomanes-Álvarez, B.R., Ibáñez, O., Navarro, F., Alemán, I., Cordón, O. and Damas, S. (2015) *Dispersion assessment in the location of facial landmarks on photographs*. International Journal of Legal Medicine, 129(1), pp.227-236. DOI:10.1007/s00414-014-1002-4

Caplova, Z., Compassi, V., Giancola, S., Gibelli, D.M., Obertová, Z., Poppa, P., Sala, R., Sforza, C. and Cattaneo, C. (2017) *Recognition of children on age-different images*. Science and Justice, 57(4), pp.250-256. DOI:10.1016/j.scijus.2017.03.005

CBC News (2002) *Mother, abducted daughter reunited after three years*. CBC. [online] Available at: http://www.cbc.ca/news/canada/mother-abducted-daughter-reunited-after-three-years-1.349507 [Accessed 11.5.2018].

CEOP (2013) *Threat assessment of child sexual exploitation and abuse*. [online] Available at: https://www.ceop.police.uk/Documents/ [Accessed 18.5.2015].

CEOP (2011) *Child trafficking update*. [online] Available at: https://www.ceop.police.uk/Documents/ [Accessed 18.5.2015].

Charman, S.D. and Carol, R.N. (2012) *Age-progressed images may harm recognition of missing children by increasing the number of plausible targets*. Journal of Applied Research in Memory and Cognition, 1(3), pp.171-178. DOI:10.1016/j.jarmac.2012.07.008

Chellappa, R., Sinha, P. and Phillips, P.J. (2010) *Face recognition by computers and humans*. IEEE, 43(2), pp.46–55.

Claes, P., Vandermeulen, D., De Greef, S., Willems, G., Clement, J.G. and Suetens, P. (2010a) *Computerized craniofacial reconstruction*. Forensic Science International, 201(1), pp.138–145.

Claes, P., Vandermeulen, D., De Greef, S., Willems, G., Clement, J.G. and Suetens, P. (2010b) *Bayesian estimation of optimal craniofacial reconstructions*. Forensic Science International, 201(1), pp.146–152.

Cui, J., Wen, F., Xiao, R., Tian, Y. and Tang, X. (2007) *Easyalbum*, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 367–376.

Dajani, E. (2008) *Intrinsic biological basis for accurate prediction of facial growth*, in: Enlow, D.H., Hans, M.G. (Eds.), Essentials of Facial Growth. Ann Arbor: Needham Press, pp. 413–424.

Dehghan, A., Ortiz, E.G., Shu, G. and Masood, S.Z. (2017) *DAGER*. arXiv:1702.04280, .

Dodge, S. and Karam, L. (2016) *Understanding how image quality affects deep neural networks*, in: 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), IEEE, pp. 1–6.

ECPAT International (2015) *Researching the Sexual Exploitation of Children: Challenges and Methodologies of Data Collection*. Bankok: ECPAT International.

ECPAT UK (2015) *The definition of trafficking: The Palermo Protocol*. [online] Available at: http://www.ecpat.org.uk/content/definition-trafficking [Accessed 22.4.2016].

Ellis, H.D., Shepherd, J.W. and Davies, G.M. (1979) *Identification of familiar and unfamiliar faces from internal and external features*. Perception, 8(4), pp.431-439.

Enlow, D.H. and Hans, M.G. (2008) *Essentials of Facial Growth*. 2nd ed. Ann Arbor: Needham Press.

Enlow, D.H. and Hans, M.G. (1996) *Essentials of Facial Growth*. 1st ed. Philadelphia: Saunders.

Erickson, W.B., Lampinen, J.M., Frowd, C.D. and Mahoney, G. (2016) *When age-progressed images are unreliable*. Science and Justice, 57(2), pp.136-143. DOI:10.1016/j.scijus.2016.11.006

European Commission (2015) *A global alliance against child sexual abuse online*. [online] Available at: http://ec.europa.eu/dgs/home-affairs/what-we-do/policies/organized-crime-and-human-trafficking/global-alliance-against-child-abuse/index_en.htm [Accessed 26.5.2015].

Facebook, Ericsson and Qualcomm (2013) *A focus on efficency*. [online] Available at: https://fbcdn-dragon-a.akamaihd.net/hphotos-ak-ash3/851560_196423357203561_929747697_n.pdf [Accessed 23.6.2015].

Farkas, L.G. (1994) *Anthropometry of the head and face*. New York: Raven Press.

Farkas, L.G., Clark, B. and Sills, D. (1994) *The use of Anthropometry in Forensic Identification*, in: Farkas, L.G. (Ed.), Anthropometry of the Head and Face. New York: Raven Press, pp. 181–190.

Farkas, L.G. and Hreczko, T. (1994) *Age-related changes in selected linear and angular measurments of the craniofacial complex in healthy North American Caucasians*, in: Farkas, L.G. (Ed.), Anthropometry of the Head and Face. New York: Raven Press, pp. 89–102.

Ferguson, E. (2015) *Facial Identification of Children: A Test of Automated Facial Recognition and Manual Facial Comparison Techniques on Juvenile Face Images*. PhD Thesis, The University of Dundee.

Fields, H.W., Proffit, W.R., Nixon, W.L., Phillips, C. and Stanek, E.D. (1984) *Facial pattern differences in long-faced children and adults*. American journal of orthodontics, 85(3), pp.217–223.

Gibson, S.J., Scandrett, C.M., Solomon, C.J., Maylin, M.I.S. and Wilkinson, C.M. (2009) *Computer assisted age progression*. Forensic Science, Medicine, and Pathology, 5(3), pp.174–181.

GOV.UK (2014) *PM announces new global action to deal with online child abuse*. [online] Available at: https://www.gov.uk/government/news/pm-announces-new-global-action-to-deal-with-online-child-abuse [Accessed 26.5.2015].

Grm, K., Štruc, V., Artiges, A., Caron, M. and Ekenel, H.K. (2017) *Strengths and weaknesses of deep learning models for face recognition against image degradations*. IET Biometrics, 7(1), pp.81–89.

Grother, P. and Ngan, M. (2015) *Child exploitation image analytics: Face recognition evaluation*. Maryland: NIST.

Grother, P. and Ngan, M. (2014) *Face Recognition Vendor Test (FRVT): Performance of face identification algorithms*. Maryland: NIST.

Grother, P.J., Ngan, M.L. and Quinn, G.W. (2017) *Face In Video Evaluation (FIVE) Face Recognition of Non-Cooperative Subjects*. Maryland: NIST.

Grother, P.J., Quinn, G.W. and Phillips, P.J. (2010) *Report on the Evaluation of 2D Still-Image Face Recognition Algorithms*. Maryland: NIST.

Hans, M.G., Broadbent, B.H. and Nelson, S.S. (1994) *The Broadbent-Bolton Growth study—past, present, and future*. American Journal of Orthodontics and Dentofacial Orthopedics, 105(6), pp.598-603. DOI:10.1016/S0889-5406(94)70145-8

Hassan, M.Y., Khalifa, O.O., Talib, A.A. and Abdulla, A.H. (2015) *Unconstrained Facial Recognition Systems*. Asian Journal of Applied Sciences, 3(2), pp.346-354.

Hennings-Yeomans, P.H., Baker, S. and Kumar, B.V.K.V. (2008a) *Simultaneous super-resolution and feature extraction for recognition of low-resolution faces*, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. DOI:10.1109/CVPR.2008.4587810

Hennings-Yeomans, P.H., Baker, S. and Kumar, B.V.K.V. (2008b) *Recognition of Low-Resolution Faces Using Multiple Still Images and Multiple Cameras*, in: 2008

IEEE Second International Conference on Biometrics: Theory, Applications and Systems, pp. 1–6. DOI:10.1109/BTAS.2008.4699321

Hofmeester (2015) *Portrait of Lotte, 0 to 16 years in 4 ½ minute*. [online] Available at: https://www.youtube.com/watch?v=-Plk7TLNmsU&t=203s [Accessed 25.4.2017].

Hofmeester (2014) *Portrait of Vince - 0 to 11 years in 4 min. (The Original)*. [online] Available at: https://www.youtube.com/watch?v=NQxyfOpZmAQ [Accessed 25.4.2017].

Hole, G.J., George, P.A., Eaves, K. and Rasek, A. (2002) *Effects of geometric distortions on face-recognition performance*. Perception, 31(10), pp.1221–1240.

Hossain, M.G., Saw, A., Ohtsuki, F., Lestrel, P.E. and Kamarul, T. (2011) *Change in facial shape in two cohorts of Japanese adult female students twenty years apart*. Singapore Medical Journal, 52(11), pp.818-823.

Huang, G.B., Mattar, M., Berg, T. and Learned-Miller, E. (2008) *Labeled faces in the wild*, in: Workshop on Faces in "Real-Life" Images: Detection, Alignment, and Recognition. Marseille, France.

Hunter, D., Tiddeman, B. and Perrett, D. (2012) *Facial ageing*, in: Caroline, W., Christopher, R. (Eds.), Craniofacial Identification. Cambridge: Cambridge University Press, pp. 57–67.

ICMEC (2013) *Confronting New Challenges in the Fight Against Child Pornography: Best practices to help file hosting and file sharing companies fight the distribution of child sexual exploitation content*. [online] Available at: https://www.icmec.org/research-library/ [Accessed 20.7.2018].

ICMEC, Carr, J. (2017) *Online child protection combatting child sexual abuse material on the internet*. [online] Available at: https://www.icmec.org/research-library/ [Accessed 20.7.2018].

INHOPE (2018) *Annual Report 2017*. [online] Available at: http://www.inhope.org/Libraries/Annual_reports/INHOPE_Annual_Report_2013.sflb.ashx [Accessed 25.7.2015].

INHOPE (2014) *Anticipate, adapt and take action*. [online] Available at: http://www.inhope.org/Libraries/Annual_reports/INHOPE_Annual_Report_2013.sflb.ashx [Accessed 25.7.2015].

International Labour Office (2008) *Combating trafficking in children for labour exploitation: A resource kit for policy-markers and practitioners*. Turin, Italy: International Labour Organization.

International Labour Office (2002) *Every child counts*. [online] Available at: http://www.ilo.org/ipecinfo/product/viewProduct.do?productId=742 [Accessed 12.5.2015].

INTERPOL (2015) *Global efforts to identify child abuse victims via INTERPOL boosted with Microsoft technology*. [online] Available at: http://www.interpol.int/News-and-media/News/2015/N2015-041 [Accessed 26.5.2015].

INTERPOL (2011) *Protecting children from actual and virtual abuse focus on INTERPOL expert meeting*. [online] Available at: http://www.interpol.int/en/News-and-media/News/2011/PR071/ [Accessed 26.5.2015].

IWF (2018) *Annual Report 2017*. [online] Available at: https://annualreport.iwf.org.uk/#new_technology [Accessed 25.7.2018].

IWF (2015a) *Emerging Patterns and Trends Report #1 Online-Produced Sexual Content*. [online] Available at: https://childhub.org/hu/child-protection-news/emerging-patterns-and-trends-report-1-online-produced-sexual-content [Accessed 20.7.2016].

IWF (2015b) *New tactics mean 137% increase in identified child sexual abuse imagery*. [online] Available at: https://www.iwf.org.uk/about-iwf/news/post/407-new-tactics-mean-137-increase-in-identified-child-sexual-abuse-imagery [Accessed 26.5.2015].

Jain, A.K., Dass, S.C. and Nandakumar, K. (2004a) *Soft biometric traits for personal recognition systems*, in: Biometric Authentication, Springer, pp. 731–738.

Jain, A.K., Ross, A. and Prabhakar, S. (2004b) *An introduction to biometric recognition*. IEEE Transactions on Circuits and Systems for Video Technology, 14 (1), pp.4–20.

John, E.A. (2017) *Track Child, Khoya-Paya web portals help trace 1.8 lakh missing children*. The Times of India. [online] Available at: https://timesofindia.indiatimes.com/india/track-child-khoya-paya-web-portals-help-trace-1-8-lakh-missing-children/articleshow/59809119.cms [Accessed 17.5.2018]

Johnston, R.A. and Edmonds, A.J. (2009) *Familiar and unfamiliar face recognition*. Memory, 17(5), pp.577-596. DOI:10.1080/09658210902976969

Karahan, S., Yildirum, M.K., Kirtac, K., Rende, F.S., Butun, G. and Ekenel, H.K. (2016) *How image degradations affect deep CNN-based face recognition?*, in: 2016 International Conference of The Biometrics Special Interest Group (BIOSIG), IEEE, pp. 1–5.

Kau, C.H. and Richmond, S. (2008) *Three-dimensional analysis of facial morphology surface changes in untreated children from 12 to 14 years of age*. American Journal of Orthodontics and Dentofacial Orthopedics, 134(6), pp.751-760. DOI:10.1016/j.ajodo.2007.01.037

Kemelmacher-Shlizerman, I., Suwajanakorn, S. and Seitz, S.M. (2014) *Illumination-aware age progression*, in: Conference On Computer Vision and Pattern Recognition (CVPR), 2014 IEEE, pp. 3334–3341.

Kong, Y., Zhang, S. and Cheng, P. (2013) *Super-resolution reconstruction face recognition based on multi-level FFD registration*. Optik - International Journal for Light and Electron Optics, 124(24), pp.6926-6931. DOI:10.1016/j.ijleo.2013.05.175

Koudelová, J., Dupej, J., Brůžek, J., Sedlak, P. and Velemínská, J. (2015) *Modelling of facial growth in Czech children based on longitudinal data*. Forensic Science International, 248, pp.33-40. DOI:10.1016/j.forsciint.2014.12.005

Kovner, A. (2018) *Thousands Of Missing Children Identified Through Facial Recognition Pilot Experiment*. IFLScience, [online] Available at: http://www.iflscience.com/technology/thousands-of-missing-children-in-india-identified-in-facial-recognition-pilot-experiment/ [Accessed 17.5.2018].

Kozbelt, A. (2001) *Artists as experts in visual cognition*. Visual Cognition, 8(6), pp.705–723.

Kuroda, S., Schmittbuhl, S. and Nanci, A. (2013) *Facial Growth and Development*, in: Nanci, A. (Ed.), Ten Cate's Oral Histology: Development, Structure, and Function. Missouri: Elsevier Inc., pp. 328–336.

Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., Von Der Malsburg, C., Wurtz, R.P. and Konen, W. (1993) *Distortion invariant object recognition in the dynamic link architecture*. IEEE Transactions on computers, 42(3), pp.300–311.

Lampinen, J., Arnal, J.D., Adams, J., Courtney, K. and Hicks, J.L. (2012a) *Forensic age progression and the search for missing children*. Psychology, Crime & Law, 18(4), pp.405–415.

Lampinen, J., Arnal, J.D., Culbertson-Faegre, A. and Sweeney, L. (2010) *Missing and Abducted Children*, in: Lampinen, J., Sexton-Radek, K. (Eds.), Protecting Children from Violence: Evidence-Based Interventions, East Sussex: Psychology Press, pp. 129–166.

Lampinen, J., Erickson, W.B., Frowd, C. and Mahoney, G. (2015) *Mighty Morphin'age progression*. Psychology, Crime and Law, 21(10), pp.952–967.

Lampinen, J., Miller, J.T. and Dehon, H. (2012b) *Depicting the Missing*. Applied Cognitive Psychology, 26(2), pp.167-173. DOI:10.1002/acp.1819

Lampinen, J. and Moore, K.N. (2016) *Missing person alerts*. Journal of Experimental Criminology, 12, pp.587-598. DOI:10.1007/s11292-016-9263-1

Lampinen, J., Peters, C.S. and Gier, V.S. (2012c) *Power in Numbers*. Applied Cognitive Psychology, 26(5), pp.702-708. DOI:10.1002/acp.2848

Lander, K., Bruce, V. and Hill, H. (2001) *Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces*. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 15(1), pp.101–116.

Lanitis, A. (2009) *Facial Biometric Templates and Aging*, in: AIAI Workshops, pp. 142–149.

Lanitis, A., Taylor, C.J. and Cootes, T.F. (2002) *Toward automatic simulation of aging effects on face images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4), pp.442–455.

Lanitis, A. and Tsapatsoulis, N. (2016) *On the analysis of factors influencing the performance of facial age progression*, in: 2016 4th International Conference on Biometrics and Forensics (IWBF), IEEE, pp. 1–6.

Lemieux, A. and Parizeau, M. (2002) *Experiments on eigenfaces robustness*, in: 16th International Conference on Pattern Recognition, 2002. Proceedings. IEEE, pp. 421–424.

Leonard, K.R. (2016) *Assessment of Facial Recognition System Performance in Realistic Operating Environments*, in: Face Recognition Across the Imaging Spectrum. Springer, Cham, pp. 117–138. DOI:10.1007/978-3-319-28501-6_6

LIAM (2015) *Age Progression and It's Reliability*. Age Progression and It's Reliability, [online] Available at: https://www.bizarrepedia.com/age-progression-reliability/ [Accessed 11.5.2018].

Liang, Y., Li, C., Yue, H. and Luo, Y. (2007) *Age simulation in young face images*, in: The 1st International Conference on Bioinformatics and Biomedical Engineering, ICBBE 2007. IEEE, pp. 494–497.

Ligthelm-Bakker, A.S., Wattel, E., Uljee, I.H. and Prahl-Andersen, B. (1992) *Vertical growth of the anterior face*. American Journal of Orthodontics and Dentofacial Orthopedics: Official Publication of the American Association of Orthodontists, Its

Constituent Societies, and the American Board of Orthodontics, 101(6), pp.509-513. DOI:10.1016/0889-5406(92)70124-S

Lin, F., Fookes, C., Chandran, V. and Sridharan, S. (2007) *Super-Resolved Faces for Improved Face Recognition from Surveillance Video*, in: Lee, S.-W., Li, S.Z. (Eds.), Advances in Biometrics, Lecture Notes in Computer Science. Presented at the International Conference on Biometrics, Springer Berlin Heidelberg, pp. 1–10. DOI:10.1007/978-3-540-74549-5_1

Lin, S.-H. (2000) *An introduction to face recognition technology*. Informing Science, 3(1), pp.1–8.

Ling, H., Soatto, S., Ramanathan, N. and Jacobs, D.W. (2010) *Face verification across age progression using discriminative methods*. IEEE Transactions on Information Forensics and Security, 5(1), pp.82–91.

Liu, C.H. and Chaudhuri, A. (2003) *Face recognition with perspective transformation*. Vision Research, 43(23), pp.2393-2402.

Liu, C.H. and Ward, J. (2006) *Face recognition in pictures is affected by perspective transformation but not by the centre of projection*. Perception, 35(12), pp.1637–1650.

Luxand, Inc. (2015) *in20years*. [online] Available at: http://www.in20years.com/ [Accessed 7.2.2015].

MacDorman, K.F. and Ishiguro, H. (2006) *The uncanny advantage of using androids in cognitive and social science research*. Interaction Studies, 7(3), pp.297–337.

Machado, C.E.P., Flores, M.R.P., Lima, L.N.C., Tinoco, R.L.R., Franco, A., Bezerra, A.C.B., Evison, M.P. and Guimarães, M.A. (2017) *A new approach for the analysis of facial growth and age estimation*. PLoS ONE, 12(7), e0180330. DOI:10.1371/journal.pone.0180330

Maheshwari, K. and Nalini, N. (2017) *Facial Recognition Enabled Smart Door Using Microsoft Face API*. International Journal of Engineering Trends and Applications, 4(3), pp.1-4.

Mahmood, Z., Ali, T. and Khan, S.U. (2016) *Effects of pose and image resolution on automatic face recognition*. IET biometrics, 5(2), pp.111–119.

Mahoney, G. and Wilkinson, C.M. (2012) *Computer-generated facial depiction*, in: Wilkinson, C.M., Rynn, C. (Eds.), Craniofacial Identification. Cambridge: Cambridge University Press, pp. 222–237. DOI:10.1017/CBO9781139049566

Malarek, V. (2002) *Two Canadian mysteries solved*. The Global and Mail. [online] Available at: https://www.theglobeandmail.com/news/national/two-canadian-mysteries-solved/article1021351/ [Accessed 14.5.2018].

Mansurov, N. (2017a) *What is Distortion?*. Photography Life. [online] Available at: https://photographylife.com/what-is-distortion [Accessed 26.1.2018].

Mansurov, N. (2017b) *Lightroom Lens Corrections Explained*. Photography Life. [online] Available at: https://photographylife.com/lightroom-lens-corrections [Accessed 26.1.2018].

Marchildon, J. (2018) *3,000 Missing Children in India Have Been Found Thanks to This New Technology*. Global Citizen. [online] Available at: https://www.globalcitizen.org/en/content/missing-children-found-india/ [Accessed 17.5.2018].

Marciniak, T., Chmielewska, A., Weychan, R., Parzych, M. and Dabrowski, A. (2015) *Influence of low resolution of images on reliability of face detection and recognition*. Multimedia Tools and Applications, 74(12), pp.4329-4349. DOI:10.1007/s11042-013-1568-8

Matthews, H., Clement, J., Kilpatrick, N., Fan, Y. and Claes, P. (2018) *Estimating age and synthesising growth in children and adolescents using 3D facial prototypes*. Forensic Science International, 286, pp.61-69. DOI:10.1016/j.forsciint.2018.02.024

Mazura, J.C., Juluru, K., Chen, J.J., Morgan, T.A., John, M. and Siegel, E.L. (2012) *Facial Recognition Software Success Rates for the Identification of 3D Surface Reconstructed Facial Images*. Journal of Digital Imaging, 25(3), pp.347-351. DOI:10.1007/s10278-011-9429-3

McGoogan, C. (2016) *Revealed: How facial recognition can expose your life to strangers*. The Telegraph. [online] Available at: https://www.telegraph.co.uk/technology/2016/04/13/revealed-how-facial-recognition-can-open-up-your-life-to-strange/ [Accessed 12.9.2018].

Mehdipour Ghazi, M. and Kemal Ekenel, H. (2016) *A comprehensive analysis of deep learning based representation for face recognition*, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 34–41.

Melendez, N. (2012) *Philadelphia man finds self on missing children's website*. CNN. [online] Available at: https://www.cnn.com/2012/04/26/us/pennsylvania-missing-mystery/index.html [Accessed 14.5.2018].

Miller, D., Kemelmacher-Shlizerman, I. and Seitz, S.M. (2015) *MegaFace: A Million Faces for Recognition at Scale*. arXiv:1505.02108.

Ministry of Woman and Child Development (2013) *National Tracking System for Missing & Vulnerable Children TrackCHILD 2.0*. [online] Available at: http://trackthemissingchild.gov.in/trackchild/index.php [Accessed 17.5.2018].

Miniwatts Marketing Group (2015) *INTERNET USAGE STATISTICS: The Internet Big Picture-World Internet Users and 2014 Population Stats*. [online] Available at: http://www.internetworldstats.com/stats.htm [Accessed 29.5.2015].

Moore, R.K. (2012) *A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena*. Scientific Reports, 2:864, pp.1-5. DOI:10.1038/srep00864

Moss, M.L. (1964) *Vertical growth of the human face*. American Journal of Orthodontics, 50(5), pp.359-376. DOI:10.1016/0002-9416(64)90176-9

MotionPortrait, Inc. (2015) *HourFace*. App Store. Available at: https://itunes.apple.com/app/hourface-3d-aging-photo/id347880622?mt=8 [Accessed 2.7.2015].

Mu, Y., Lo, H., Ding, W. and Tao, D. (2014) *Face Recognition from Multiple Images Per Subject*, in: Proceedings of the 22Nd ACM International Conference on Multimedia, MM '14. New York, NY, USA: ACM, pp. 889–892. DOI:10.1145/2647868.2655054

Mullins, J. (2012) *Age progression and regression*, in: Caroline, W. and Christopher, R. (Eds.), Craniofacial Identification. Cambridge: Cambridge University Press, pp. 68–75.

Nanda, S.K. (1988) *Patterns of vertical growth in the face*. American Journal of Orthodontics and Dentofacial Orthopedics: Official Publication of the American Association of Orthodontists, Its Constituent Societies, and the American Board of Orthodontics, 93, pp.103–116.

Natu, V. and O'Toole, A.J. (2011) *The neural processing of familiar and unfamiliar faces*. British Journal of Psychology, 102(4), pp.726-747. DOI:10.1111/j.2044-8295.2011.02053.x

NCMEC (2016) *Long-Term Missing Child Guide for Law Enforcement*. [online] Available at: http://www.missingkids.org/publications/longtermmissingguide [Accessed 13.3.2017].

NCMEC (2015a) *Annual Report of the National Center for Missing & Exploited Children*. [online] Available at: http://www.missingkids.org/AnnualReport [Accessed 13.3.2017].

NCMEC (2015b) *Key facts*. [online] Available at: http://www.missingkids.com/KeyFacts [Accessed 18.5.2015].

NDTV (2018) *Facial Recognition System Helps Trace 3,000 Missing Children In 4 Days*. NDTV.com, [online] Available at: https://www.ndtv.com/india-news/facial-recognition-system-helps-trace-3-000-missing-children-in-4-days-1841192 [Accessed 17.5.2018].

Ochoa, B.K. and Nanda, R.S. (2004) *Comparison of maxillary and mandibular growth*. American Journal of Orthodontics and Dentofacial Orthopedics, 125(2), 148–159. DOI:10.1016/S0889540603008540

OHCHR (2000) *Protocol to Prevent, Suppress and Punish Trafficking in Persons especially Women and Children, supplementing the*. [online] Available at: http://www.ohchr.org/EN/ProfessionalInterest/Pages/ProtocolTraffickingInPersons.aspx [Accessed 22.4.2016].

O'Toole, A.J., Vetter, T. and Blanz, V. (1999) *Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition*. Vision Research, 39(18), pp.3145–3155.

Panis, G. and Lanitis, A. (2014) *An Overview of Research Activities in Facial Age Estimation Using the FG-NET Aging Database*, in: Agapito, L., Bronstein, M.M. and Rother, C. (Eds.), Computer Vision - ECCV 2014 Workshops, Lecture Notes in Computer Science. Springer International Publishing, pp. 737–750.

Panis, G., Lanitis, A., Tsapatsoulis, N., Cootes, T.F. (2016). *Overview of research on facial ageing using the FG-NET ageing database*. IET Biom. 5, pp.37–46. DOI:10.1049/iet-bmt.2014.0053

Parkhi, O.M., Vedaldi, A. and Zisserman, A. (2015) *Deep face recognition*. Proceedings of the British Machine Vision, 1(3), 1–6. DOI:10.5244/C.29.41

Phillips, P.J., Yates, A.N., Hu, Y., Hahn, C.A., Noyes, E., Jackson, K., Cavazos, J.G., Jeckeln, G., Ranjan, R. and Sankaranarayanan, S. (2018) *Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms*. Proceedings of the National Academy of Sciences, 115:24, pp.6171-6176.

Protalinski, E. (2016) *Google is shutting down Picasa on May 1, 2016*. [online] Available at: https://venturebeat.com/2016/02/12/google-is-shutting-down-picasa-on-may-1-2016/ [Accessed 15.9.2017].

Qin, C., Bao, X., Roy Choudhury, R. and Nelakuditi, S. (2011) *TagSense*, in: Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services, MobiSys ’11. New York, NY, USA: ACM, pp. 1–14. DOI:10.1145/1999995.1999997

Rafferty, Y. (2007) *Children for sale*. Child Abuse Review, 16(6), pp.401-422. DOI:10.1002/car.1009

Raji, J.M., Garba, S.H., Numan, A.I., Waziri, M.A. and Maina, M.B. (2010) *Morphological evaluation of head and face shapes in a North-Eastern Nigerian population*. Australian Journal of Basic and Applied Sciences, 4(8), pp.3338–3341.

Rakosi, T., Jonas, I. and Graber, T.M. (1993) *Orthodontic Diagnosis*. New York: Thieme Medical Publishers.

Ramanathan, N. and Chellappa, R. (2006) *Modeling age progression in young faces*, in: 2006 Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE, pp. 387–394.

Ranjan, R., Sankaranarayanan, S., Castillo, C.D. and Chellappa, R. (2017) *An all-in-one convolutional neural network for face analysis*, in: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 17–24.

Rawat, W. and Wang, Z. (2017) *Deep convolutional neural networks for image classification*. Neural computation, 29(9), pp.2352–2449.

Reid, D., Samangooei, S., Chen, C., Nixon, M. and Ross, A. (2013) *Soft biometrics for surveillance*. Handbook of Statistics, 31, pp.327–352. DOI:10.1016/B978-0-444-53859-8.00013-8

Reid, D.A. and Nixon, M.S. (2011) *Using comparative human descriptions for soft biometrics*, in: 2011 International Joint Conference On Biometrics (IJCB). IEEE, pp. 1–6.

Ronneburger, A., Basarab, J. and Howland, H.C. (2006) *Growth of the cornea from infancy to adolescence*. Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians (Optometrists), 26(1), pp.80-87. DOI:10.1111/j.1475-1313.2005.00362.x

Sabharwal, A. (2016) *Moving on from Picasa*. Moving on from Picasa, [online] Available at: http://googlephotos.blogspot.co.uk/2016/02/moving-on-from-picasa.html [Accessed 15.9.2017].

Sassouni, V. and Nanda, S. (1964) *Analysis of dentofacial vertical proportions*. American Journal of Orthodontics, 50(11), pp.801-823. DOI:10.1016/0002-9416(64)90039-9

Scherbaum, K., Sunkel, M., Seidel, H.-P. and Blanz, V. (2007) *Prediction of individual non-linear aging trajectories of faces*, in: Computer Graphics Forum. Wiley Online Library, pp. 285–294.

Schroff, F., Kalenichenko, D. and Philbin, J. (2015) *Facenet*. eprint arXiv:1503.03832. DOI:10.1109/CVPR.2015.7298682

Seyama, J. and Nagayama, R.S. (2007) *The Uncanny Valley*. Presence: Teleoperators and Virtual Environments, 16(4), pp.337-351. DOI:10.1162/pres.16.4.337

Shah, S., Khanal, L., Koirala, S. and Bhattacharya, S.C. (2015) *The study of Prosopic index of 17-26 year old normal population in Eastern Nepal*. Rushian Open Medical Journal, 4(2), 1–3. DOI:10.15275/rusomj.2015.0201

Shen, C.-T., Huang, F., Lu, W.-H., Shih, S.-W. and Liao, H.-Y.M. (2014) *3D Age Progression Prediction in Children's Faces with a Small Exemplar-Image Set.* Journal of Information Science and Engineering, 30(4), pp.1131–1148.

Silva, R.F., Pereira, S.D., Prado, F.B., Daruge, E. and Daruge, E. (2008) *Forensic odontology identification using smile photograph analysis–case reports*. Journal of Forensic Odonto-Stomatology, 26(1), pp.12–17.

Sinha, P., Balas, B., Ostrovsky, Y. and Russell, R. (2006) *Face recognition by humans*. Proceedings of the IEEE, 94(11), pp.1948–1962.

Smith, S. (2014a) *Briefing Paper – Preliminary Analysis of New Commercial CSAM Website Accepting Payment by Bitcoin*. [online] Available at: https://www.iwf.org.uk/sites/default/files/inline-files/Preliminary_analysis_into_commercial_CSAM_distributor_accepting_bitcoin_sanitised_not_restricted_01014.pdf [Accessed 25.7.2018].

Smith, S. (2014b) *Briefing Paper - Rogue Affiliates Distributing CSAM using "Disguised Websites" (Public version)*. [online] Available at: https://www.iwf.org.uk/sites/default/files/inline-files/Analysis_of_rogue_affiliates_commercial_public_0414.pdf [Accessed 25.7.2018].

Statistic Brain Research Institute (2015a) *Social Networking Statistics*. [online] Available at: http://www.statisticbrain.com/social-networking-statistics/ [Accessed 29.6.2015].

Statistic Brain Research Institute (2015b) *Facebook Statistics*. [online] Available at: http://www.statisticbrain.com/facebook-statistics/ [Accessed 29.6.2015].

Stephan, C.N. (2015) *Perspective distortion in craniofacial superimposition*. Forensic Science International, 257, 520.e1-520.e8. DOI:10.1016/j.forsciint.2015.09.009

Stone, Z., Zickler, T. and Darrell, T. (2010) *Toward Large-Scale Face Recognition Using Social Network Context*. Proceedings of the IEEE, 98(8), pp.1408-1415. DOI:10.1109/JPROC.2010.2044551

Taigman, Y., Yang, M., Ranzato, M. and Wolf, L. (2014) *Deepface*, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 1701–1708.

Taylor, K.T. (2000) *Forensic Art and Illustration*. Florida: CRC Press.

Thomee, B., Shamma, D.A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D. and Li, L.-J. (2016) *YFCC100M*. Communications of the ACM, 59(2), pp.64–73. DOI:10.1145/2812802

Torres-Restrepo, A.M., Quintero-Monsalve, A.M., Giraldo-Mira, J.F., Rueda, Z.V., Vélez-Trujillo, N. and Botero-Mariaca, P. (2014) *Agreement between cranial and facial classification through clinical observation and anthropometric measurement among envigado school children*. BMC Oral Health, 14:50, 1–8. DOI:10.1186/1472-6831-14-50

Toseeb, U., Keeble, D.R.T. and Bryant, E.J. (2012) *The Significance of Hair for Face Recognition*. PLoS ONE, 7(3), e34144. DOI:10.1371/journal.pone.0034144

Turk, M. and Pentland, A. (1991) *Eigenfaces for recognition*. Journal of Cognitive Neuroscience, 3(1), pp.71–86.

UNODC (2016) *Global Report on Trafficking in Persons*. [online] Available at: https://www.unodc.org/unodc/en/data-and-analysis/glotip.html [Accessed 8.2.2018].

UNODC (2015) *Human Trafficking*. [online] Available at: http://www.unodc.org/unodc/en/human-trafficking/what-is-human-trafficking.html [Accessed 7.3.2015].

UNODC (2014) *Global Report on Trafficking in Persons*. [online] Available at: https://www.unodc.org/unodc/en/data-and-analysis/glotip_2014.html [Accessed 8.2.2018].

UNODC (2004) *United Nations Convention Against Transnational Organized Crime and the Protocols Thereto*. [online] Available at: https://www.unodc.org/unodc/en/organized-crime/intro/UNTOC.html [Accessed 12.5.2015].

U.S. Department of Justice (2010) *The National Strategy for Child Exploitation Prevention and Interdiction*. [online] Available at: https://www.justice.gov/psc/publications [Accessed 18.5.2015].

U.S. Department of State (2012) *Country Reports on Human Rights Practices for 2012*. [online] Available at: https://www.state.gov/j/drl/rls/hrrpt/2012/ [Accessed 13.5.2015].

U.S. Department of State (2007) *Trafficking in Persons Report*. [online] Available at: https://www.state.gov/j/tip/rls/tiprpt/ [Accessed 18.5.2015].

VGT (2011) *Who we are – Virtual Global Taskforce*. [online] Available at: http://www.virtualglobaltaskforce.com/who-we-are/ [Accessed 28.5.2015].

Vogt, S., Magnussen, S. (2007) *Expertise in pictorial perception*. Perception, 36, pp.91–100.

Wagner, M. (2015) *Ohio teen applying for college discovers he was abducted from Alabama home in 2002*. New York Daily News. [online] Available at: /news/crime/alabama-boy-abducted-2002-found-alive-ohio-article-1.2424344 [Accessed 14.5.2018].

Wang, J., Zhang, C. and Shum, H.-Y. (2004) *Face image resolution versus face recognition performance based on two global methods*, in: Proceedings of Asia Conference on Computer Vision, pp. 48–49.

Wei, W. (2012) *Online Child Sexual Abuse Content: The Development of a Comprehensive, Transferable International Internet Notice and Takedown System*. [online] Available at: https://www.iwf.org.uk/resources/independent-report-on-international-notice-and-takedown-system [Accessed 25.7.2018].

Wheeler, F.W., Liu, X. and Tu, P.H. (2007) *Multi-Frame Super-Resolution for Face Recognition*, in: 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1–6. DOI:10.1109/BTAS.2007.4401949

Whitefoot, J. (2018) *What Happened to Jaycee Dugard, Kidnapping Survivor and Mother?*. Crimeola. [online] Available at: https://crimeola.com/jaycee-dugard-kidnapping-survivor-mother/ [Accessed 11.5.2018].

Wu, H.-J., Chung, C.Y., Shi, X. and Normile, J. (2013) *Facial pose improvement with perspective distortion correction*. US 8,599,238 B2.

Wu, T., Chellappa, R. (2012) *Age Invariant Face Verification with Relative Craniofacial Growth Model*, in: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y. and Schmid, C. (Eds.), Computer Vision – ECCV 2012, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 58–71.

Yao, D. (2014) *Baobeihuijia.com Helps 804 Families Find Missing Children*. [online] Available at: http://www.fuzhouexpat.com/baobeihuijia-com-a-website-helped-804-families-reunited/ [Accessed 18.5.2015].

Yi, D., Lei, Z., Liao, S. and Li, S.Z. (2014) *Learning face representation from scratch*. eprint arXiv:1411.7923.

Zhu, C., Wen, F. and Sun, J. (2011) *A rank-order distance based clustering algorithm for face tagging*, in: Computer Vision and Pattern Recognition (CVPR) 2011. IEEE, pp. 481–488.

Zhu, X. and Ramanan, D. (2012) *Face detection, pose estimation, and landmark localization in the wild*, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2879–2886.

Acknowledgements

Figure 6:

© 2006 IEEE. Reprinted, with permission, from Ramanathan, N. and Chellappa, R., Modeling age progression in young faces, 2006 Computer Society Conference on Computer Vision and Pattern Recognition, 2006.

Figure 7:

© 2014 IEEE. Reprinted, with permission, from Kemelmacher-Shlizerman, I., Suwajanakorn, S. and Seitz, S.M., Illumination-aware age progression, 2014 IEEE Conference on Computer Vision and Pattern Recognition, June, 2014.

Figure 8:

Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, On Facial Age Progression Based on Modified Active Appearance Models with Face Texture by Bukar A.M., Ugail H., Hussain N. In: Advances in Computational Intelligence Systems by Angelov P., Gegov A., Jayne C., Shen Q. (eds). © 2017.

Figure 9:

From Scherbaum, K., Sunkel, M., Seidel, H.-P. and Blanz, V. Prediction of individual non-linear aging trajectories of faces, in: Computer Graphics Forum, pp.285-294. Copyright © 2007 by John Wiley Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.

Figure 11:

Reprinted from Forensic Science International, 248, Koudelová, J., Dupej, J., Brůžek, J., Sedlak, P. and Velemínská, J, Modelling of facial growth in Czech children based on longitudinal data, pp.33-40, © 2015, with permission from Elsevier.

Figure 12:

Reprinted from Forensic Science International, 248, Matthews, H., Clement, J., Kilpatrick, N., Fan, Y. and Claes, P., Estimating age and synthesising growth in children and adolescents using 3D facial prototypes, pp.61-69, © 2018, with permission from Elsevier.

Figure 13:
Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Forensic Science, Medicine, and Pathology, Computer assisted age progression, Gibson, S.J., Scandrett, C.M., Solomon, C.J., Maylin, M.I.S. and Wilkinson, C.M. © 2009.

Figure 14:
Reprinted from Science & Justice, 57(2), Erickson, W.B., Lampinen, J.M., Frowd, C.D. and Mahoney, G., When age-progressed images are unreliable, pp.136-143, © 2017, with permission from Elsevier.

Figure 17:
Copyright: © 2017 Machado et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Figure 20:
© Torres-Restrepo et al.; licensee BioMed Central Ltd. 2014
This article is published under license to BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

Figure 21:
Reprinted from American Journal of Orthodontics and Dentofacial Orthopedics, 93(2), Nanda, S.K., Patterns of vertical growth in the face, pp.103-116, © 1988, with permission from Elsevier.